AMRL-TDR-63-87

HUMAN GROUP PERFORMANCE DURING CONFINEMENT

EARL A. ALLUISI, Ph.D.
LOCKHEED-GEORGIA COMPANY

W. DEAN CHILES, Ph.D.
BEHAVIORAL SCIENCES LABORATORY
6570th AEROSPACE MEDICAL RESEARCH LABORATORIES

THOMAS J. HALL, M.S.I.E.
LOCKHEED-GEORGIA COMPANY

and

GLENN R. HAWKES, Ph.D.
LOCKHEED-GEORGIA COMPANY

TECHNICAL DOCUMENTARY REPORT No. AMRL-TDR-63-87

NOVEMBER 1963

6570th AEROSPACE MEDICAL RESEARCH LABORATORIES
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

Contract Monitor: W. Dean Chiles, Ph.D.
Project No. 1710, Task No. 171002

(Prepared in Part under Contract No. AF 33(616)-7607-M4 by
Human Factors Research Laboratory, Research Directorate
Lockheed-Georgia Company, Marietta, Georgia)

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Qualified requesters may obtain copies from the Defense Documentation Center (DDC), Cameron Station, Alexandria, Virginia. Orders will be expedited if placed through the librarian or other person designated to request documents from DDC formerly ASTIA).

Do not return this copy. Retain or destroy.

Stock quantities available at Office of Technical Services, Department of Commerce, Washington 25, D. C. Price per copy is $2.00.

Change of Address

Organizations receiving reports via the 6570th Aerospace Medical Research Laboratories automatic mailing lists should submit the addressograph plate stamp on the report envelope or refer to the code number when corresponding about change of address.

800 - December 1963 - 162-21-345

FOREWORD

This report was prepared by E. A. Alluisi, T. J. Hall, and G. R. Hawkes of the Lockheed-Georgia Company and by W. D. Chiles of the 6570th Aerospace Medical Research Laboratories. The research reported here was begun in December 1961 and was completed at the end of November 1962.

The two studies described in the report were conducted in the Human Factors Research Laboratory, Office of the Director of Research, Lockheed-Georgia Company, under the direction of Dr. Earl A. Alluisi, Project Director, with the assistance of Thomas J. Hall, Glenn R. Hawkes, George D. Hayes, James N. Howard, H. Douglas Meyers, William J. Sonenshine, and Herman E. Williams.

The work was supported by the Behavioral Sciences Laboratory, 6570th Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, under Contract No. AF 33(616)-7607-M4, Project No. 1710, "Training Personnel and Psychological Stress Aspects of Bioastronautics," and Task No. 171002, "Performance Effects of Environmental Stress." Dr. W. Dean Chiles, Assistant Chief, Training Research Division, served as task scientist and consultant.

The assistance of a number of persons, whose services were obtained to supplement the laboratory staff especially during the data-collection and monitoring phases of the experimentation, is gratefully acknowledged. Included among these extra workers at the experimenter's station were about twenty undergraduate and graduate students of psychology from the Georgia Institute of Technology, the University of Georgia, Emory University, Georgia State College, and Kansas State University. Although it would not be practical to list all their names here, these students performed their assigned tasks with the responsibility, intelligence, and maturity that clearly credits their institutions. Mr. F. P. Carter of the Slater System provided outstanding food servicing to an extent, and of a quality seldom experienced for prolonged periods; to him, and to the other personnel of the Slater System, both the experimenters and the subjects wish to have their gratitude expressed. Finally, the authors wish to thank Dr. Oscar S. Adams, Operations Research Division, Lockheed-Georgia Company, for his substantial contributions to every phase of the research.

ABSTRACT

Six Air Force Academy cadets were confined for 15 days in a simulated advanced-system crew compartment while following a schedule of 4-hours on duty and 2-hours off, and two 5-man crews of USAF pilots were confined for 30 days while alternating shifts on a schedule of 4-hours on duty and 4-hours off. While on duty the operators were tested with a battery of 6 performance tasks, 2 of which required interactions among crewmembers in the form of exchanges of information, cooperation, and temporal coordination. In addition, the data of the present studies were compared with those of two previous 15-day tests of two crews who worked the 4-2 schedule while being tested with a battery of 5 individual-performance tasks.

The data suggest that with proper control of selection and motivational factors, crews can work effectively for periods of at least 2 weeks and probably longer using a schedule of 4-hours on duty and 2-hours off. Crews can work even more effectively for periods of at least a month and quite probably for 2 or 3 months using a schedule of 4-hours on duty and 4-hours off, and with this schedule less demanding controls of selection and motivational factors are required.

PUBLICATION REVIEW

This technical documentary report is approved.

Walter F. Grether

WALTER F. GRETHER
Technical Director
Behavioral Sciences Laboratory

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (cont'd)

LIST OF TABLES

HUMAN GROUP PERFORMANCE DURING CONFINEMENT

INTRODUCTION

An important factor to be considered in planning for a multiple-man space-vehicle system is the extent to which the mission will be affected by the crew's ability to perform under unusual and potentially stressful work-environment conditions. A crew will probably consist of a small number of men who will be confined to their vehicle for periods of several weeks or months. The crew compartment will be severely limited in volume to avoid prohibitively high requirements for rocket power. These few crewmembers will have to maintain high-level, around-the-clock system performance; to achieve this level of system performance without increasing the crew size to unacceptable limits, optimum scheduling of work and rest will be necessary. An ideal schedule would maximize the working period efficiently, while creating neither deleterious physiological effects nor decrements in performance. Similar requirements for high-level, around-the-clock, alert performance and optimum work scheduling exist in other advanced systems — indeed, even in some earthbound systems where logistical considerations dictate that the number of crewmembers assigned per operator position be minimized insofar as practicable (e.g., DEW Line and BMEWS sites).

During the past several years, the Lockheed-Georgia Company has developed special laboratory facilities and conducted a number of studies of both individual and crew performance under the contractual support and scientific monitoring of the 6570th Aerospace Medical Research Laboratories. In the earliest experimental study (reported in detail by Adams and Chiles, ref. 2), the performance of 16 subjects was measured over a period of 96 hours on four different duty-period and rest-period schedules (2 hours of work and 2 hours of rest, 4 work and 4 rest, 6 and 6, and 8 and 8). The results of the study suggested that for both active tasks (mental arithmetic and pattern perception) and passive tasks (monitoring and vigilance), the 2-hour and 4-hour shifts were superior to the others, particularly on the bases of subject preference and over-all indices of subject adjustment to the test periods. It was evident that the subjects could work at the tasks assigned with maintained efficiency for 12 hours per day over periods at least as long as 96 hours.

Two additional 96-hour experiments followed the initial study in order to examine greater proportions of time at work. In one of these the subjects followed a schedule of 4 hours of work and 2 hours of rest, and in the other they followed a schedule of 6 hours of work and 2 hours of rest. Both of these studies have been reported in full elsewhere (Appendix I of Adams and Chiles, ref. 3). The performance data obtained in these studies did not show that either schedule was superior, but evidence obtained from questionnaires completed by the subjects suggested that severe decrements in performance would probably have resulted from prolongation of the 6-2 schedule beyond the 96 hours of testing. For example, the subjects on the 6-2 schedule averaged less than 4 hours of sleep per day, whereas those on the 4-2 schedule averaged 5.5 or more hours of sleep per day. Unless the conditions of space flight reduce sleep requirements (a possibility not yet demonstrated), 4 hours of sleep per day are considered to be inadequate over prolonged periods of time (cf. Ray, Martin, and Alluisi, ref. 6).

These 4-day experiments were followed by a long-term investigation, "Operation 360," in which two crews of operational personnel were separately confined to a relatively small (1100 cubic feet) crew compartment and tested on a 4-2 schedule over 15-day periods (Adams and Chiles, ref. 3). During this period the subjects had no communications with the "outside world," and communications with the experimenters were limited to intercom messages of direct relevance to the "mission." The 4-2 schedule was selected because it showed promise of being the most efficient and practicable of the schedules previously studied. When working such a schedule, three men with appropriate cross-training can operate two positions continuously. The 15-day period of confinement was selected because such a length of time extends beyond the point at which individuals can be expected to compensate by extra effort for serious degrees of fatigue-induced deterioration; also, the choice was influenced by the belief that the subjects would consider 15 days to be a long time — and, therefore, not a trivial experience — if the situation were to have proved to be unpleasant. Practical considerations of economy, equipment reliability, and maintenance (at the time the study was conducted) also acted to preclude the use of a longer period of confinement.

Well marked diurnal rhythms were evidenced in the levels of both performance efficiency and autonomic activation; these hourly variations equaled or exceeded the variations between daily means in most cases. Subsequently a control study was carried out in which 10 college students worked 4 hours per day for 5 consecutive days per week over a 6-week period. The trends in performance of the confined subjects as compared with the control group led to the conclusion that with a minimum of selection subjects could be found whose motivation and abilities would lead to acceptable levels of performance on a schedule of 4-hours work and 2-hours rest for periods at least as long as 15 days.

These data were encouraging, but they evidenced two shortcomings. First, they suffered from being based entirely on individual, as compared with group-dependent performance. Second, they apply specifically to missions of durations no greater than 15 days. It was in order to collect data not subject to these restrictions that the research reported here was conducted.

Two group-performance tasks were devised to provide quantitative data relative to group-dependent performances; these tasks were then tested (Alluisi, Hall, and Chiles, ref. 5), and some modifications and improvements made in them before use in the present studies. In each of these two tasks, successful performance requires interactions among crewmembers in the form of exchanges of information, cooperation, and coordination. One of the tasks, "code-lock solving," stresses temporal coordination. The other group-performance task is an elaboration of an individual task of "target identification." This task, which is highly dependent on individual levels of proficiency, requires the filtering of visual information from simulated "noisy" displays of sensed targets. Both tasks require whole-crew participation and interaction for successful performance, and both provide quantitative indices of crew performance. They are described in detail in a later section of this report.

Two experimental investigations have been completed and are described here. In the first, "HOPE-II," a 6-man crew worked for 15 days using the 4-2 schedule; thus, the data of this study can be compared in some respects with those of the previous 15-day experiment, "Operation 360" (Adams and Chiles, ref. 3), in which only individual performance tasks were used. In the second study, "HOPE-III," two 5-man crews were tested over a 30-day confinement period; the crews worked a less demanding schedule of 4-hours on duty and 4-hours off in alternate shifts throughout the 30 days. In addition, in order to avoid the typical "end effect" that has appeared as a "spurt" in performance during the last two or three days of many long-term confinement studies, the subjects in the 30-day study were led to expect a 40-day confinement. They did not know that the study was to end short of the 40-day period until the mockup door was opened at the end of the 30th day and they were addressed by the task scientist.

The data obtained under the conditions of this false expectation can be projected to periods of performance beyond 30 days more validly than would have been the case had the subjects known that the study would end at 30 days. In addition to precluding the anticipated "end spurt" or "end effect," the 40-day duration of expected confinement meant that the subjects would feel there was still a long time to go right up to the end of the study. In other words, "sticking it out" would thereby be a factor right up to the end of the study. Any difficulties that might have occurred during the 31st through the 40th days could be inferred to be chance phenomena if no problems had developed with the subjects during 30 days of confinement. This same conclusion was apparently reached independently by the subjects, for all of them informally agreed with it in unsolicited statements at the end of the study; they reported that in their opinions nothing would have been gained by them (beyond that of which they had already been convinced) by extension of the study to an actual 40-day period of confinement.

Questions of logistical and environmental support were not being investigated in this or in the previous studies of this series. All of these studies have been predicated on the assumption that those scientists who have responsibility for such support areas will be successful in achieving their goals, and that they will then be able to simulate or provide in space the same sort of relatively comfortable and life-sustaining environment used here.

METHOD

SUBJECTS

The subject sample for the 15-day study, HOPE-II, consisted of six Air Force Academy Cadets who volunteered to serve in the tests. These subjects, all of whom were physically qualified for flight training, were selected from among approximately 75 volunteers; all volunteers understood that participation in the study would eliminate 20 days of their annual 30-day period of leave. The six subjects ranged in age from 19 to 22 years, with a median age of 21; one of the six celebrated his 21st birthday while confined during the study. On entering the mockup, the subjects ranged in weights between 138 and 178 pounds, with a median of 159 pounds; one subject lost 3 pounds during confinement, but all others gained weight so that the terminal weights ranged from 144 to 185, with a median of 162 pounds.

The subject sample for the 30-day study, HOPE-III, consisted of 10 Air Force officers who volunteered to serve in the tests. The 10, all of whom were rated Air Force Pilots, were selected from among 36 volunteers of the Air Training Command, Undergraduate Pilot Training Class 62-H. The criteria used in selecting the subjects from among the volunteers were, in broad outline, those used in the selection of Project Mercury Astronauts. The selection was made by personnel of the Air Surgeon's staff at Headquarters, Air Training Command. One crew of five men had been in training together at Reese AF Base, Texas, whereas the other crew had been in pilot training together at Vance AF Base, Oklahoma. The two crews were comparable in all other respects. The 10 subjects ranged in age from 23 to 26 years, with a median of 24; 2 subjects, one in each of the HOPE-III crews, celebrated their 24th birthdays while confined. On entering the mockup, the subjects ranged in weight between 158 and 204 pounds, with a median of 190 pounds; most subjects actively dieted during confinement, and as a result six lost weight, two remained constant, and only two gained weight. The terminal weights ranged from 165 to 198, with a median of 188 pounds.

TEST FACILITY

Subjects were tested in an advanced-system crew-compartment mockup consisting of three sections: a 5-station work area, a leisure area, and a sleeping area. The total volume is approximately 1100 cubic feet, half of which is devoted to the work-station and leisure areas, and half to the sleeping area (a detailed description of the work-station and leisure areas is reported in ref. 1).

While on duty, each subject occupied an assigned position in the work station area. During off-duty hours subjects were restricted to the leisure area and the adjacent 6-bunk sleeping area. The work-station and leisure areas were adequately illuminated at all times. The sleeping area was maintained in a semi-darkened condition, the sole illumination being provided by two low-intensity light sources located at floor level. A system of small speakers located at different points throughout the mockup was used to present broad-band noise continuously at a constant level of about 85 decibels (Sound Pressure Level) in the work-station and leisure areas and about 70 decibels (SPL) in the sleeping area. The noise was used for purposes of realistic simulation and to mask all extraneous outside sounds. Audiograms, taken on all subjects before and after their 15- or 30-day exposures, revealed no evidence of hearing losses attributable to these experiments.

PERFORMANCE AND PSYCHOPHYSIOLOGICAL MEASURES

Major Performance Tasks — A battery of six tasks designed to test both individual and small group performance provided the principal behavioral measures in this study. The tasks were displayed on each of five performance panels (one at each operator station). The face of one of these panels is shown in figure 1. Four of these tasks were selected

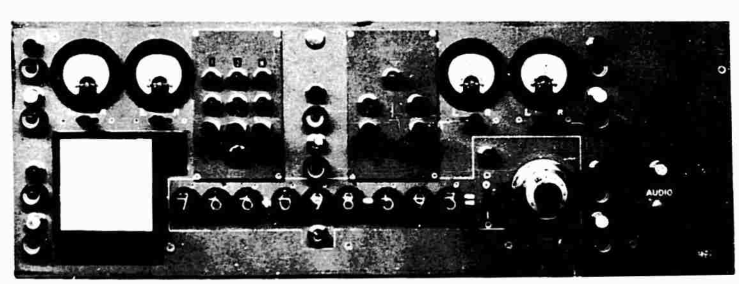

Figure 1. Front view of the performance-task panel.

initially on the basis of an analysis of individual operator requirements for long-range, long-endurance weapon systems (cf. Adams, ref. 1). They have been modified slightly during the several years of use, and the two group-performance tasks have been added recently. All tasks show very high reliabilities (see Alluisi, Hall, and Chiles, ref. 5) and have done so since their earliest use (see Adams, Levine, and Chiles, ref. 4). The four individual tasks are arithmetic computation, probability monitoring, warning-lights monitoring, and auditory vigilance. The two group-performance tasks are code-lock solving* and target identification. Only the target identification task will be described in detail here since it was introduced for the first time (in its present form) in these studies; the other tasks have been retained essentially as described fully in previous reports (refs. 2, 3, and 5).

At the lower left corner of each performance panel (see figure 1), there was a 4-inch square, 6 x 6 matrix of 36 lights. Various combinations of these lights could be illuminated to present patterns that gave the appearance of vertically oriented, solid bargraphs. The population of possible figures was constrained by the rule that each of the six possible column heights appeared once and only once in any given figure. Thus, the total number of different possible figures was 6!, or 720. Of this number, 240 figures were drawn at random to serve as the standard "target images." Comparison images were drawn at random from the same population.

The task was presented to the subject in the following manner. First, one of the target images appeared for a period of 5 seconds followed by a 5-second off period. Then a comparison image was presented for 2 seconds followed by a 2-second off period. Next a second comparison image was presented for 2 seconds. This was followed by a 14-second response period during which the subject was required to indicate whether the first, the second, or neither comparison image was the same as the target image.

The 240 target images were arranged in the same order on each of 15 punched tapes, but the correct answer associated with a given target image (first, second, or neither) was random and different on each tape. Also, different comparison images were used on each tape for presentations of those images that were not the same as the target image. Within each subgroup of 60 problems (i.e., the first, second, third, and fourth quarters of each tape of 240 target images and 480 comparison images) the correct solutions were equally divided among the three response categories (20 each first, second, and neither).

The task was complicated further by the introduction of random distortions of the comparison images. Although the target image was undistorted, each of the comparison images was perturbed with visual "noise" that was generated through use of a secondary noise program — a program that was separate and distinct from the basic punched tape program used to generate the target and the comparison images. This secondary program was designed to operate so that the instruction sent by the basic program to any given light on a given subject's display could be reversed. For example, assume that the secondary program designated a given light to be perturbed (i.e., to be noisy) on a given presentation. If the basic program were to instruct that light to be lit on that trial, the secondary program would prevent the light from coming on; if the basic program were to instruct the light to be off, the secondary program would turn it on.

Six different levels of visual noise were used by perturbing 3, 4, 5, 6, 8, or 10 lights on the display at a given operator position. The number of noisy lights at a given position remained constant for a 24-hour period, but the specific lights that were noisy varied randomly from problem to problem throughout the study. In HOPE-II, the levels

*The code-lock task requires the subjects to discover the correct sequence in which each of five buttons (one at each operator position) has to be pushed to illuminate a green light and to do this as quickly as possible without neglecting other duties. Since only four men were on duty at any given time in HOPE-II, the task was modified for that study by providing a second code-lock button to one of the four subjects on duty. Thus each of the three subjects who operated "alpha" and "bravo" positions in HOPE-II was required from time to time to operate two buttons on the code-lock task.

of visual noise were arranged so that two of the four crewmembers on duty at a given time would be exposed to a high-noise level and the other two would be exposed to a low-noise level. During the first 5 days of the study, 3 of the lights (16.7%) were noisy for the low-noise condition and 6 (33.3%) for the high condition. During the second 5-day block, the noise levels were 27.8% (5 lights) and 55.6% (10 lights). During the final 5 days, the noise levels were 33.3% and 44.4% (8 lights).

In HOPE-III, 3 members of the 5-man crew on duty received a high-noise condition and the remaining 2 received a low condition. The actual noise levels employed during the 5-day blocks of testing were as follows: block-1, 22.2% and 33.3%; block-2, 27.8% and 44.4%; block-3, 33.3% and 44.4%; blocks-4, -5, and -6 replicated blocks-1, -2, and -3, respectively. The assignment of noise levels to crew positions within each 5-day block was varied systematically so that a given subject would not be assigned a low-noise level two days in a row, and would be confronted with a high-noise level no more than 2 days in a row.

The group aspect of the task was introduced by presenting each subject's decision to the crew commander on an auxiliary display. The commander was then required to enter a final decision based on the responses made by each operator (including himself); this final decision as to whether first, second, or neither comparison image was the same as the target image was defined as a command responsibility. The choice made by the crew commander in this was displayed to each subject by the illumination of one of three green cue lights on each operator panel. Immediately before the programmer moved on to the next problem, one of three blue lights was illuminated on the crew commander's auxiliary panel to indicate the correct answer to the problem. The commander was free to inform the crew, if he so chose, of the correct answer to the problem or the correctness of the final decision; from this information, each crewmember could infer whether or not his individual answer was correct.

Since this task had not been used previously, decisions with respect to the noise levels to be used had to be reached on the basis of less extensive pilot experimentation. One single value could have been selected for the low-noise level (and one for the high), and this could have been used throughout HOPE-II. However, had this been done, and had the choice been an unfortunate one (and this could have occurred because the pilot experiments necessarily involved less practice on the task than provided during the full course of confinement), it was possible that meaningful data would not be obtained concerning the effects of duration of confinement nor about the nature of the task itself. On the other hand, the study of target-identification performance over the total duration of confinement could be partially sacrificed in favor of the use of several different levels of high and low noise — a technique that would be more likely to yield information about the task, but information about performance changes only within the 5-day blocks of confinement. HOPE-III followed too closely on the heels of HOPE-II to permit detailed data analyses that were adequate to select fixed noise levels for the later study. Thus it was that the noise levels were varied over the 5-day blocks of both studies as described above.

Subject Logs — Each subject was provided with a log book in which he was asked to record throughout the study comments pertinent to his reactions, feelings, etc. In addition, the subjects in HOPE-II were instructed to include a specific type of comment during each 2-hour work period. The instructions concerning these special comments were approximately as follows: "Think of a possible problem that might be encountered in an aerospace vehicle mission. Devise some solution to that problem, and enter both the problem and the solution in the lower half of each page in your log. You should enter one such problem and solution on each page, i.e., one for each 2-hour period of the study." Inspection of the "problem-solution" comments made by HOPE-II led to the elimination of this requirement for HOPE-III.

Psychophysiological Measures – Two psychophysiological measures were selected for use in each of the two experiments. One measure was pulse rate, and the other was axillary temperature. Both measures were taken by each subject on himself, and the self-determined measures were recorded by the subject on the appropriate page of his own log. In HOPE-II, each subject was required to enter the measures on himself on each page of the log: i.e., for each 2-hour period of testing. He was required to take the measures within 15 minutes of the even hours of the day. In HOPE-III the same requirements held, except that no subject was required to interrupt his sleep in order to obtain or record this information (in HOPE-III the subjects had 4-hour off-duty periods, whereas they had only 2-hour off-duty periods in HOPE-II). The temperature and pulse records were kept faithfully by all subjects in complete accord with their instructions.

GENERAL PROCEDURE

The 6 subjects comprising the HOPE-II crew were tested on an around-the-clock schedule of 4 hours of duty and 2 hours of rest for a period of 15 days. For example, on a typical day, subjects 1, 3, 4, and 6 would be on duty from 0800 to 1000 hours, and subjects 2 and 5 would be off duty. At 1000 hours, subjects 3 and 6 would be replaced by subjects 2 and 5, and at 1200 hours subjects 3 and 6 would return to duty to replace subjects 1 and 4. With the 6-man crew used, this schedule made it possible to keep four work stations occupied continuously except for a very brief time (less than a minute) during which the replacements took place. At the end of the 15 days, each subject had accumulated a total of 240 hours of work in HOPE-II.

The 10 subjects comprising the HOPE-III crews were tested on an around-the-clock schedule of 4 hours of duty and 4 hours of rest for a period of 30 days. These subjects were organized into two crews of 5 men each, and the crews worked alternating 4-hour shifts. For example, on a typical day, ABLE crew (consisting of subjects 1, 3, 5, 7, and 9) would be on duty from 0800 to 1200 hours, and BAKER crew (consisting of subjects 2, 4, 6, 8 and 10) would be off duty. At 1200 hours, BAKER crew would replace ABLE, and would remain on duty until relieved at 1600 hours. Thus, all five work stations in the mockup were occupied continuously except for the brief interval of time required for the change of crews. At the end of the 30-day testing period, each of the 10 subjects had accumulated a total of 360 hours of work in HOPE-III.

TASK PROGRAM

While on duty, the subjects in both experiments worked according to a basic 2-hour task program. This program, which included scheduled periods that placed low, intermediate, and high demands on performance, was designed to be as compatible as possible with the programs previously used (cf. refs. 2, 3, and 5). The program is shown in table 1; it was repeated 180 times during the 15-day tests, and 360 times during the 30-day tests.

As indicated in the table, the program consisted of a 30-minute low-demand performance period, 60 minutes of intermediate-demand performance, and 30 minutes of high-demand performance during each 2-hour period of testing. From the subject's point of view there was no break between repetitions of the program since the three passive tasks (auditory vigilance, warning-lights monitoring, and probability monitoring) were presented continuously at each work station. An amber light on each panel signaled that the arithmetic task would begin in 30 seconds; another amber light provided a similar 30-second warning for target identification. The subjects were told that their performance was being scored continuously; however, analysis has been made only of the data obtained during the 90-minute intermediate- and high-performance periods. Also, it should be noted that the terms "low," "intermediate," and "high" are purely relative in the sense used here. The absolute levels of demand implied by these terms are not likely to have remained constant; rather, they are more likely to have changed over the course of confinement, for example as functions of continued practice.

Table 1

Basic 2-Hour Task-Performance Schedule

Task	Minutes 00	15	30	45	60	75	90	105	120
Auditory Vigilance	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	
Warning-Lights Monitoring	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	
Probability Monitoring	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	
Arithmetic Computation		xxx	xxx						
Code-Lock Solving			xxx	xxx	xxx	xxx			
Target Identification						xxx	xxx		
Level of Demand on Performance	low	med	hi	med	med	hi	med	low	

NOTE: Each "x" in the table represents 5 minutes.

ORIENTATION AND TRAINING

The crews for HOPE-II and HOPE-III reported to the laboratory to begin orientation and training on their respective reporting dates; in both cases this was scheduled for 0800 hours on a Thursday. After official welcomes, most of the first day was devoted to orientation, general briefings concerning the purpose of the research projects, and descriptions of the test plans; the second day was spent in training, interviews, and physical examinations.

In general, during the orientation and training periods considerable time was spent in discussing with the subjects the objectives of the study and the importance to the Air Force of obtaining quantitative data relating to the problem of optimum work-rest schedules. In addition, about 20 minutes extra were spent with the HOPE-II crew in showing them the diurnal variations in performance and physiological activation evidenced in the previous 15-day confinement studies (cf. Adams and Chiles, ref. 3), and in emphasizing to them that one of the principal purposes of the HOPE-II tests was to determine to what extent the diurnal differences in performance could be changed simply by informing the subjects of the fact that such differences would occur unless they did something about it. The same point was made to the HOPE-III crews, but in their case the point was not emphasized as a principal purpose of the study; rather, it was presented as a standard operating procedure under which the subjects were instructed to attempt to minimize diurnal variations in performance by expending extra effort during periods when they felt poorly.

Physical conditions in the laboratory area were arranged so that the subjects could interact with only the Air Force task scientist and the Lockheed project leader during the orientation phase. While training in the mockup, the subjects could interact only with the three experimenters who served as shift leaders, and then only over the intercom. In other words, a cordial but semiformal and business-like relation was established and maintained between the crewmembers and the experimenters at all times. Questions and comments were encouraged, and every attempt was made to dispel any doubt, fear, or sense of mystery that may have developed concerning the nature of the tests and the period of confinement.

Pretest interviews were held privately with the individual subjects to provide opportunities for the expression of any anxieties concerning participation in the study (none was evidenced) and, further, to obtain information about items such as the subject's age, marital status, nearest relative, anticipation of possible family or personal emergencies, and specific sources of worry. At the same time, each subject was given, if he wanted them, a supply of postcards on which he could write messages to his family or friends. These cards and letters were collected by the experimenter together with instructions from the subject concerning the dates on which they should be posted. The subjects were assured that any mail which arrived for them during the period of confinement would be held safely for delivery at the time of termination of confinement.

The rapport between the experimenters and subjects was excellent both during training and throughout the experiment. This was confirmed by experimenter observations, voluntary comments of subjects, and the responses of subjects to questions asked during post-test debriefing interviews.

TESTING

Each crew reported to the laboratory at about 0700 hours on a Saturday and spent nearly an hour in the laboratory before being confined for the beginning of the study at exactly 0800 hours. During this time the crewmembers ate breakfast, stowed their personal equipment in individual lockers located in the mockup, and were given a final pretest briefing during which any last-minute questions were answered. Reading material was permitted in the mockup, provided it was taken in prior to confinement; although all subjects were observed to have stowed some reading material, the only appreciable amount of reading done was by subjects in the 30-day tests whose 4-4 work-rest schedule was less demanding than the 4-2 work-rest schedule of the subjects in the 15-day tests.

Incidental to the actual performance testing, the subjects in both groups had been given detailed instructions concerning the following regulations and procedures related to the test conditions:

(a) A clearly audible horn was sounded in the sleeping quarters 10 minutes prior to the time a change in shift was scheduled, except that twice daily in HOPE-III the horn was sounded 10 minutes prior to a scheduled mealtime (when the off-duty crew was scheduled to eat a meal before going on duty). In order to stop the horn it was necessary for one man to get out of his bunk and actuate a toggle switch. The 10-minute interval allowed sufficient time for the subjects to dress and wash. Exactly on the even-hour marks in HOPE-II (and at alternate even-hour marks in HOPE-III), when subjects were scheduled to change shifts. the experimenter sounded three short blasts on the wake-up horn as a signal for the change of stations.

(b) Meals were placed in a food compartment in the leisure area of the mockup at scheduled mealtimes. The meals were selected from the daily plant cafeteria menus by the chief experimenter. Throughout the 15-day HOPE-II tests, the subjects were permitted to eat during the 30-minute low-performance portions in the middle of their 4-hour work periods. At the proper time, when the low-performance period began, one of the two subjects scheduled for a meal would be designated by the crew commander to leave his work station, remove the food trays from the food compartment, and bring them to the work-station area. When the meals had been eaten, a subject returned the trays to a garbage-disposal shoot in the leisure area. The subjects were instructed to continue working on the monitoring and vigilance tasks even while eating.

(c) During the 30-minute low-performance period of the task program, subjects were allowed to leave their duty stations to use the toilet facility if they were given permission to do so by the crew commander. Only one on-duty crewmember was permitted to be away from his station at any given time.

(d) Standard operating procedures were established for use of the intercom system. In general, standard radio procedure was followed. The test crew of the 15-day experiment was designated "HOPE-II". The two test crews of the 30-day experiment were designated "HOPE-III" with "ABLE" and "BAKER" being used to designate each specific crew. The experimenter station was designated "GELAC CONTROL." The subjects were instructed to keep intercom conversation to a minimum, and the only calls permitted between the subjects and GELAC CONTROL were "business" calls such as those required to report an apparent malfunction of equipment. Any calls initiated by the subjects that were other than business in nature were acknowledged and then cordially, but firmly, discouraged. Whenever GELAC CONTROL called the crew, the call was addressed to the crew commander; if for some reason another member of the crew was to be addressed, the crew commander was contacted first and his permission was requested to address a specific member of the crew. The major intercom system was a "common-line" system in which all stations (including the experimenter station) received all communications. Two additional intercom systems were provided; one of these, a "private-line" system, connected a speaker and microphone in the leisure area with the experimenter station, whereas the second system (connecting the same two points) used telephone-type receivers to provide a "very private" line. These latter two systems were installed for use in the case of personal emergencies, medical consultation, etc.

(e) The subjects were informed that with the beginning of the confinement period the test would be considered a closed-system operation. Their only contact with the outside would be with the experimenter by means of the intercom system. They were assured, however, that any given subject would be released from the test in case of illness or in the case of emergencies at home.

(f) The food service unit located in the leisure area was well stocked with supplemental food items, including canned fruit juices, soups, instant coffee, tea, milk, crackers, and an assortment of gum, candy, and nuts. These could be consumed ad libitum. Cigars, cigarettes, and pipe tobacco were provided with meals according to choices (and quantities) indicated by the subjects prior to confinement.

On the day following their termination of confinement, each subject was interviewed, given a physical examination, and was asked to complete the questionnaires and a test of interpersonal knowledge scheduled for the post-test period. In addition, all the subjects were interviewed privately to obtain their reactions to a series of questions relating to attitudes toward the experiment and experimenters, adjustments to the work schedules, opinions as to task difficulties, and any other subjective reactions that they wished to voice.

RESULTS

SUBJECT LOGS

Subjects in both tests faithfully recorded their self-determined temperature and pulse rates in their logs as instructed; these psychophysiological measures are analyzed and discussed in a subsequent section. The subjects in HOPE-II were widely variable in their responses to the request that during each 2-hour period they list a problem and its solution. The number of problems and solutions actually reported by each of the 6 subjects ranged from 45 to 180, with a median of 112. Because of the incompleteness of these data, and because the subjects indicated they had great difficulty in thinking up new and novel problems, this requirement was not imposed on the HOPE-III subjects.

In both studies, extensive and regular voluntary comments were made only by the crew commanders. Apparently the crew commanders felt responsible for recording observations on their crews' performance efficiency, morale, and motivation. They also frequently suggested various minor modifications of the tasks and the physical facilities of the mockup. There appeared to be no discrepancies either among the comments of the crew commanders or between those of the crew commanders and the crewmembers; the only noticeable difference was that the crew commanders made comments with greater frequency and in greater detail than crewmembers.

The topic of sleep was the most frequent subject of the comments. Specific comments were made regarding amount of sleep obtained, adjustments to loss of sleep, and thoughts concerning the sleep patterns necessary to maintain a stable level of performance on the atypical schedules used. In general, the comments suggested that all subjects had difficulty initially (a) in falling asleep quickly enough to use efficiently the off-duty periods, and (b) in maintaining alertness with less than the normal amount of sleep - "normal" being uniformly considered 8 hours per 24-hour period. Several HOPE-III crewmembers reported apparent shifts in diurnal high and low periods of alertness during the study; the shifts were reported to be tendencies toward greater alertness during the early evening hours as contrasted with initial feelings of maximum alertness only during the daytime hours. Little or no evidence for such shifts in alertness was obtained from the psychophysiological and performance measures taken during the experiment; where averaged over the members of both crews, only the axillary-temperature data even suggested that such a shift might have occurred at a psychophysiological level.

According to the comments entered in the logs, most of the crewmembers made satisfactory adjustments to the necessity for falling asleep quickly by the end of the first few days of confinement. They appeared also to have adjusted to other aspects of the 4-2 or 4-4 work-rest schedules they followed; for example, most subjects reported that they had adjusted to fewer than 8 hours of sleep per day, and reported further that 6 hours of sleep appeared sufficient for them to maintain alertness and a sense of physical well-being.

Notations concerning the performance tasks were made frequently by the crew commanders and also (but not as frequently) by the other crewmembers. Most of the comments concerned adjustments to the 2-hour performance-task program and thoughts regarding methods for maximizing performance efficiency on specific tasks. The crew commanders also recorded regularly any and all departures from the established duty or off-duty routines. Even minor equipment repairs, such as the replacement of a light bulb in an indicator light, were reported in some detail.

All subjects made note of minor physical discomforts. The most frequently mentioned irritations were those associated with feelings of soreness in those parts of the body that had physical contact with the environment. For example, several subjects mentioned chafing of the elbows, sore buttocks, and sore spots on the head and ears; all of the comments were understandable reactions to changes made necessary by the atypical aspects of the confinement, especially those that required the subjects to sit at work stations and wear headphones while working 16 hours per day in HOPE-II and 12 in HOPE-III. Occasionally, cold symptoms, headaches, foot irritations, and sensations of burning of the eyes were reported; all of these conditions were relieved or ameliorated by self-application of appropriate minor medicants from the medicine chest in the mockup. Some subjects made note of elimination difficulties which they were able to overcome principally by the ingestion of special food items such as prune juice.

Recorded comments concerning the food indicated that all the subjects considered it to be of excellent quality. Although equal quantities were provided each subject in both experiments, the subjects in HOPE-II believed they were given too little food, whereas those in HOPE-III thought they were supplied too much. These beliefs apparently led the crews to differential regulation of caloric intake with HOPE-II subjects supplementing their meals with the candies, soups, and juices provided, and the HOPE-III subjects consciously dieting (many keeping a continuous calorie count for each 24-hour period along with a record of weight self-determined each day by use of scales available in the leisure area).

The crew commanders commented frequently on the apparent morale and level of motivation of their crews and the individuals comprising them. According to these reports, morale started off at a high level, but tended to drop somewhat during the first few days of confinement. Morale then remained relatively constant except for noticeable rises at several specific points; namely, (a) at the midpoints of the expected confinement periods, or the 7th and 8th days in HOPE-II and the 20th day in HOPE-III, (b) during the last two or three days of HOPE-II when the subjects knew the study was about to end, and (c) during the celebration of a crewmember's birthday - an experience shared by the members of each crew. During these latter occasions a birthday cake was provided and a congratulatory message was called in via intercom to the appropriate crew commander with the request that he convey the message to the individual concerned. In addition, the crew was permitted to leave its duty station for 20 minutes, during a 30-minute low-performance period, in order to celebrate the birthday and consume the cake.

A few notations were made concerning irritability among crewmembers. These were made principally during the early portions of confinement. The incidents were ascribed generally to sleepiness and fatigue. None of the incidents was thought to be serious by the crewmembers; apparently, none adversely affected task performance. The most frequent comment concerning morale and motivation in general related to the boredom felt by the subjects (apparently universally) once the task program had been learned well enough for it to seem routine. This, of course, is a condition that the program was expected to create, and it is a condition that causes problems in the sense of making it difficult to maintain motivation at high levels. It is a condition that is widespread in actual systems where many of the activities required of the operator are repetitive in nature.

The subjects in HOPE-II engaged in essentially no off-duty activities other than those related to personal hygiene and sleep. However, the comments of the HOPE-III subjects indicated that considerable amounts of time were devoted to reading, letter writing, and games. Of the games engaged in, chess apparently became the early favorite, but then lost favor because the mental excitement it stimulated in the players tended to interfere with their sleep. Subsequently, the games that became the more favored (such as dominoes) were those that required little sustained mental activity and did not persist in thought after termination of the game.

Only rarely were entries of a personal nature made in the subject logs. Where these did occur, they were generally in the form of negative comments regarding one or another of the crewmembers and his recent behavior in the work situation, especially on the crew-performance tasks.

The subjects were told prior to entering the confinement that it was relatively easy to hear "voices, music, etc." in the white noise, and some such auditory illusions were reported, though only very rarely. Also, some subjects reported occasionally having very vivid dreams. None of the subjects indicated any concern about either of these kinds of experiences.

HOPE-II ACTIVITY CHECKLIST AND EXPERIMENTER LOGS

The experimenters made systematic observations of each of the on-duty crewmembers at least once every 15 minutes throughout the 15-day HOPE-II tests. The observations were recorded on an "Activity Checklist" that was used as an aid in standardizing both the method and the tabulation of the data. In addition to keeping the Activity Checklist, the experimenters maintained a log in which they recorded the salient points of any extended intercom conversations between any of the crewmembers and "GELAC CONTROL."

The Activity Checklist provided seven major categories under which ratings were made: (a) GENERAL APPEARANCE, (b) ATTENTION TO ACTIVE TASKS, (c) ATTENTION TO PASSIVE TASKS, (d) POSTURE, (e) CURTAIN (referring to the curtain located at each duty station to screen it from all others), (f) HEADSET, and (g) REMARKS. The ratings were highly consistent, with over 83% of the responses indicating "highly alert" appearance, "continuously responding" attention to active tasks, "systematically scanning" attention to passive tasks, "slouching" posture, "closed" curtain, and headset "on." Since the ratings appeared not to be related to performance, the Activity Checklist was not used in HOPE-III.

The crewmembers of HOPE-II made no observed attempts to regulate caloric intake. Although all of the subjects performed some exercises in the leisure area during off-duty hours, the crew commander apparently did not establish this as a routine procedure with a specified schedule. The subjects made several requests during the period of the test for various food items and for certain nonprescription medicines. The criterion used in deciding whether to honor these requests or not was as follows: The request would be honored if it was judged that the item requested would have been stored aboard the mockup prior to the experiment had the experimenters been given sufficient time to become fully familiar with the subjects; otherwise, the request would not be honored. The nonprescription medicines requested were for treatment of common cold symptoms, mouthwash, and a lotion for acne.

Two more-serious medical problems arose during the 15-day confinement. Each of these required consultations and examinations by a physician. Examinations could be made by means of a closed-circuit television system that operated through a one-way vision glass in the ceiling of the mockup, and consultations between the physician and his patient were completed by use of the very private intercom system previously mentioned. One of the problems was created by the loss of a temporary cap for a tooth that had recently undergone a root-canal operation. In this case the physician decided in consultation with the subject that definitive dental treatment could await completion of the confinement period; he supplied the subject with appropriate medicants for self-administered treatment during the test, and arrangements were made for final treatment by an oral surgeon on the day the subject was released from confinement.

The second medical problem was associated with one subject's complaint of a progressively more painful earache. On the fourteenth day of the study the physician felt it necessary to make a brief (about five minutes) examination of the infected ear. This was done by having the other off-duty subject remain in the leisure area while the ill subject was escorted through an emergency exit in the sleeping area to a room immediately adjacent to the mockup. The subject had contact only with the physician and the chief experimenter during the examination. Examination by the physician revealed small boils in the external meatus of the right ear, and treatment by ingestion of antibiotic tablets was prescribed. During the treatment period, moreover, the physician made regular checks of the subject's condition, and the crew commander was asked to observe and report systematically his observations of the subject's performance. This subject was permitted to have one additional 2-hour off-duty period. This gave him the chance to obtain about 4 hours of sleep to make up for sleep losses incurred during other off-duty periods because of the earache. The treatment was apparently successful, for by the end of the test the irritation had disappeared and the general condition of the ear was much improved.

HOPE-III EXPERIMENTER LOGS

Many of the entries in the HOPE-III experimenter logs were concerned with equipment difficulties and the extent to which these interfered with data collection; typically, these were of a minor nature (e.g., burned-out light bulbs). When a malfunction was reported by any of the subjects, a system check was run to determine the nature of the trouble, and another check was run after repair to ensure that the equipment had been put back in good working order. (Over the duration of the entire 30-day tests, all equipment functioned normally more than 99% of the time).

On two occasions the off-duty crew overslept; i.e., the alarm was sounded at the proper time, and it was turned off by someone in the crew, but no one arose. Both of these occasions occurred with the crew scheduled to arise and eat breakfast before going on duty, and in both cases the crew then on duty aroused the relief crew in time for the proper change of shift at 0800 hours. The late-sleeping crew, on these two occasions, ate breakfast during the 15-minute low-performance section of the program that immediately followed their movement into the work stations. The only other entries in the experimenter logs concerning off-duty activities were (a) that a game of chess was in progress during several periods, and (b) that lengthy discussions concerning politics and the world situation took place on several other occasions.

As was the case in the 15-day confinement tests, the subjects in this 30-day confinement made several requests for replenishment of the stored stocks of fruit juices, soups, cigaretts, paper cups, etc. In addition, they requested non-prescription medications for ills such as sore throats, colds, athelete's foot, and fever blisters. On two occasions, a physician conversed with individual subjects and prescribed treatment for mouth ulcer and hordeolum. Requests made by the subjects during the period of confinement were judged with the same criterion used in the HOPE-II tests, and in general all requests were filled.

PSYCHOPHYSIOLOGICAL MEASURES

The methods used in obtaining the pulse and temperature data in HOPE-II provided a measure for each subject on each measure within a quarter of an hour of 0800, 1000, 1200, etc. for each of the 180 even hours of the study; those of HOPE-III provided similar measures for 3 out of every 4 of the 360 even hours of the 30 days, the fourth period being one during which the subject was asleep. During each 24-hour period of "Operation 360" (OPN-360), the 15-day study that preceded HOPE-II (Adams and Chiles, ref. 3), the psychophysiological measures were recorded 4 times for each subject at intervals of 4 and 8 hours, alternately, while the subject was performing arithmetic computations. The data from all three of these studies have been analyzed in terms of the means of measurements taken on different days (between-day trends) and in terms of the means of measurements made at different times of day (within-day trends). Comparisons have been made throughout, where appropriate and possible, among the data of HOPE-II, HOPE-III, and OPN-360. In order to obviate making assumptions as to the forms of the sampling distributions, nonparametric statistical tests have been used here and wherever possible in analyzing the results.

Between-Day Trends - The psychophysiological data of the three groups have been summarized slightly differently, as necessitated by the differences in the times and numbers of measurements. Since two scorings of each subject in OPN-360 were available for each 12-hour interval, "rolling-mean" scores were used to represent each of 12 different 2-hour intervals per day. Thus, the ordinate value of each point in the between-day psychophysiological scores of OPN-360 constitutes the average of the 22 individual scores (2 from each of 11 subjects) obtained during the 3 preceding and 3 following, immediately adjacent, 2-hour experimental periods; the abscissa value is represented as the average of the six appropriate period midpoints.

Similarly, rolling means were used in plotting the data of HOPE-II; these means were taken over 3 basic data points so as to include each subject 3 times — once before starting to work, once in the low-performance period between 2 work periods, and once after having completed 4 hours of work at his station. With the data of the HOPE-III subjects, rolling means were taken over 4 basic data points (i.e, over 4 successive and immediately adjacent 2-hour experimental periods); thus, each subject is represented 3 times in each plotted point (as in HOPE-II), with the missing fourth point representing a time during which the subject was asleep.

The mean pulse-rate data of OPN-360 and HOPE-II, and of HOPE-III are presented in figures 2 and 3, respectively.* The corresponding mean axillary temperatures are presented in figures 4 and 5. In order to show conveniently the relation of between-day trends to within-day trends, the 15-day curves of OPN-360 and HOPE-II have been divided into three 5-day segments and the 30-day curves of HOPE-III have been divided into six 5-day segments. Also, in order to assist in making direct comparisons, the data of OPN-360 and HOPE-II have been plotted on the same figure; likewise, the two 15-day segments of the 30-day HOPE-III data have been plotted together.

Friedman's analysis-of-variance test for ranked data (Chi-square-r) was used to test for the significance of between-day effects. Each subject's daily mean was assigned a rank between 1 and 15 (for OPN-360 and HOPE-II), or between 1 and 30 (for HOPE-III), and these ranks were used in determining the level of significance of any over-all trends that were evidenced by the data. The results of these tests are given in table 2. Also given in the table are the separate results for each of the two 5-man crews that served in HOPE-III; the fact that essentially identical trends held for the 2 crews served as the justification for combining the data of the 2 crews into the single curves of figures 3 and 5.

---

* Figures 2 and 3, and all subsequent figures are grouped for convenience at the end of this report, pages 40 to 72.

Table 2

Summary of Analysis of Variance by Ranks:
Daily Levels of Psychophysiological Measures

	OPN-360*		HOPE-II		HOPE-III	
Measure	Chi-square	P less than**	Chi-square	P less than**	Chi-square	P less than**
Pulse Rate	57.11	.001	13.80	---	93.78	.001
					A = 51.70	.01
					B = 67.62	.001
Temperature	10.46	---	23.21	---	63.55	.001
					A = 76.55	.001
					B = 50.57	.01

N.B. Self-determined pulse rates and axillary temperatures were data of HOPE-II and HOPE-III. Sampled heart-rate levels and skin-temperature levels were data of OPN-360. Under HOPE-III, "A" and "B" are the two 5-man crews.

* Adopted from Adams and Chiles (ref. 3).
** Based on df of 14 for OPN-360 and HOPE-II, and on df of 29 for HOPE-III.

In terms of pulse rate, it is apparent from figure 2 that the subjects of OPN-360 began their 15-day confinement with a relatively fast pulse rate that declined from about the fifth day until a stable level was reached on about the tenth day. The HOPE-II subjects, on the other hand, began their 15-day confinement with a relatively slow pulse rate that began climbing on about the fourth day and continued climbing until a relatively stable level was reached on about the twelfth day. As indicated in table 2, the over-all trend was statistically significant in the case of OPN-360, but not in the case of the HOPE-II subjects. The significant trend in the over-all pulse rates of the HOPE-III subjects was, like that of the OPN-360 one of a declining pulse rate; the decline, however, was of lesser extent.

The between-day trends in the temperature measurements of OPN-360 and HOPE-II were neither consistent (see figure 4) nor statistically significant (see table 2). In contrast, for HOPE-III there was a small, but consistent and statistically significant lowering of axillary temperature (see figure 5 and table 2).

The analyses summarized in table 2 of daily levels of the psychophysiological measures reflect the consistencies in temporal trends among subjects; an indication of the consistencies within each of the subjects is given in the data of table 3. The rank-order coefficients of correlation (rho's) between the daily means and the days of the study are shown in table 3 for each subject. A positive correlation represents an increase in the measure with continuation of the study, a negative correlation represents a decrease.

Table 3

Rank-Order Coefficient of Correlation (Rho) of Each Subject's Daily Psychophysiological Measures With Days of the Studies

Group	Subject	Psychophysiological Measure Pulse	Temperature
OPN-360 *	1A	-.75	(-.29)
	2A	-.91	(-.17)
	3A	-.78	(.31)
	4A	-.62	(.41)
	5A	-.72	-.52
	1B	-.45	-.44
	2B	(-.02)	-.65
	3B	(-.16)	(.19)
	4B	(-.22)	.45
	5B	-.84	(.01)
	6B	-.48	.75
HOPE-II **	1	.62	(.28)
	2	(-.23)	.68
	3	.67	.47
	4	.92	.90
	5	.84	(.42)
	6	-.83	(-.41)
HOPE-III ***	1A	-.60	(.00)
	3A	(-.16)	-.45
	5A	-.61	(-.14)
	7A	(-.23)	.34
	9A	(.10)	(.19)
	2B	-.62	-.63
	4B	-.90	-.51
	6B	(-.12)	(.29)
	8B	-.43	.37
	10B	-.42	(.20)

* Heart-rate level and skin-temperature level were the measures used in OPN-360; adopted from Adams and Chiles (ref. 3). N = 15 in each case, so for a one-tailed test rho = .44 is associated with P = .05, and rho = .62 is associated with P = .01.

** Self-determined pulse and axillary temperature. N = 15, significant values of rho are identical to those of OPN-360.

*** Self-determined pulse and axillary temperature. N = 30, so for a one-tailed test, rho = .31 is associated with P = .05, and rho = .43 is associated with P = .01.

N.B. Parentheses are used to enclose nonsignificant values of rho.

Among the 11 subjects in OPN-360, 8 showed significant negative trends in pulse rate and none showed a significant positive trend (3 had nonsignificant trends); 3 indicated significant negative trends in temperature, 2 indicated significant positive trends, and 6 showed no significant trends. Four of the 6 subjects in HOPE-II had significant positive trends in pulse, 1 had a significant negative trend, and the remaining 1 had essentially no trend; 3 of these same subjects had significant positive trends in temperature, and 3 indicated no significant trends. The pulse-rate trends of 6 of the 10 subjects in HOPE-III were negative, and there were no significant positive trends; on the other hand, in terms of temperature, 2 subjects demonstrated significant positive trends, 3 significant negative trends, and 5 no significant trends at all.

Within-Day Trends - One of the most prominent features of the pulse-rate and temperature data presented in figures 2 through 5 is the reasonably consistent within-day change, or diurnal variation, in these measures. Indeed, it is apparent that these variations continued throughout the 15-day confinements of OPN-360 and HOPE-II, as well as the 30-day confinement of HOPE-III. In the case of the latter, however, there is some indication of reduced variations (or flattenings of the curves) in both measures towards the end of the study (viz., during days 26 to 30). These variations are shown more clearly in figures 6 through 9 where the respective measures have been combined for the different hours of the day on the basis of the three 5-day blocks of experimentation in OPN-360 and HOPE-II, and the six 5-day blocks in HOPE-III.

Friedman's test was used to determine the significance of the observed differences among the within-day means for each of the 5-day blocks and for the entire period of experimentation. The results of these analyses are summarized in table 4 for HOPE-II and HOPE-III. Because only 4 points were available per day for each of the subjects in the OPN-360, and because these 4 points occurred at different times of day, those data were not subjected to statistical analysis. However, the consistency with which the within-day changes occurred among the members of OPN-360 strongly suggests that, for the subjects tested, the changes are beyond chance expectation (cf. Adams and Chiles, ref. 3, p 23).

Table 4
Summary of Analysis of Variance by Ranks:
Within-Day Levels of Psychophysiological Measures

		HOPE-II		HOPE-III	
Measure	Day Blocks	Chi-square	P less than*	Chi-square	P less than*
Pulse Rate	1-5	21.14	.05	38.17	.001
	6-10	15.02	---	41.95	.001
	11-15	16.70	---	40.96	.001
	16-20	n.a.	n.a.	34.79	.001
	21-25	n.a.	n.a.	21.38	.01
	26-30	n.a.	n.a.	21.91	.01
	1-15 or 1-30	19.52	---	40.29	.001
Axillary Temperature	1-5	24.44	.02	53.61	.001
	6-10	20.91	.05	45.65	.001
	11-15	22.97	.02	37.82	.001
	16-20	n.a.	n.a.	34.31	.001
	21-25	n.a.	n.a.	19.53	.02
	26-30	n.a.	n.a.	11.85	---
	1-15 or 1-30	29.07	.005	40.19	.001

* Based on _df_ of 11 for HOPE-II, and on _df_ of 8 for HOPE-III. n.a. — not applicable

A statistically significant within-day effect is present in both the pulse-rate and temperature data of HOPE-III when averaged over the entire 30-day period of confinement; however, there is not only a noticeable flattening of the diurnal variations toward the end of the 30 days, (see figures 5 and 9) but also a failure of the temperature differences to achieve statistical significance during the last 5-day block of experimentation. There is no noticeable flattening of HOPE-II's axillary temperature curves over the three 5-day blocks of experimentation, and the diurnal variations within each of the blocks are statistically significant, as are also the variations taken over the total 15-day period. The within-day changes in HOPE-II's pulse rate are significant only for the first 5-day block, and the over-all 15-day averages are not significant.

If pulse rate and temperature are taken as indices of the level of physiological activation, it may be noted that there were apparently shifts in the time at which the highest activation occurred with the subjects over the course of the confinements. The OPN-360 subjects evidenced highest activation during the early evening hours (about 1800 hours) of the first 5 days, but during the late night hours (about 2200 hours) of the third 5-day block of experimentation. An identical shift of 4 hours is evidenced in the data of the HOPE-II subjects for axillary temperature, but the shift in pulse rate is from 1600 hours to 2000 hours (see figures 6 and 8).

Pulse rate for the subjects of HOPE-III (figure 7) shows at most a 2-hour shift of the point of highest activation (from 1800 to 2000 hours) over the first 15 days, and a shift back (from 2000 to 1800 hours) over the last 15 days, for a net zero change over the 30-day period. Axillary temperature, however, shifted in a manner more nearly consistent with the data of the other two groups; as shown in figure 9, the point of maximum activation of the HOPE-III subjects shifted over the first 15 days by 2 hours (from 1800 to 2000 hours) and it shifted again over the final 15 days by the same amount and in the same direction (from 2000 to 2200 hours).

PERFORMANCE MEASURES, INDIVIDUAL TASKS

The results obtained on the individual-performance aspects of 5 of the 6 performance tasks in the battery (the code-lock task yields group-performance scores only) have been analyzed both in terms of the means of measurements made on the different days (between-day trends) and in terms of the means of measurements made at different times of day (within-day trends). As was the case with the psychophysiological measures, comparisons have been made wherever possible between the data of the HOPE-II and HOPE-III performance and the comparable data of the previously reported OPN-360 tests. Also, non-parametric tests have been used to obviate the making of any assumptions concerning the forms of the sampling distributions of the performance measures.

Between-Day Trends - The mean levels of performance attained on the individual tasks are presented in figures 10 through 19. The data of OPN-360 and HOPE-II are results of the 120, 2-hour periods worked by the subjects in those studies; the data of HOPE-III are results of the 180, 2-hour periods worked by each subject in the 30-day study.

Because of the arrangement of the duty periods used with the 4-2 schedule of OPN-360 and HOPE-II, a measure that averaged the results of a 6-hour time span had to be used if each subject were to make an equal contribution to each data point. For example, from 1000 to 1200, subjects 1, 2, 3, and 4 would be on duty; from 1200 to 1400, subjects 3, 4, 5, and 6 would be on duty; from 1400 to 1600 subjects 1, 2, 5, and 6 would be on duty. Thus, if the basic 2-hour period were used as the analysis unit (e.g., 1000 to 1200) two subjects would not contribute to the performance measure (for the first period in the example, subjects 5 and 6 would not be included). If a 4-hour period were used as the unit (e.g., 1000 to 1400), subjects 3 and 4 would contribute 4 hours of performance, whereas subjects 1, 2, 5, and 6 would each contribute only 2 hours. Over a six-hour period (e.g., 1000 to 1600), each subject would contribute 4 hours of performance.

Without modification, however, use of this technique would yield only four data points per day. In order to avoid this and to obtain a data point for each 2-hour period, a rolling-mean technique was used. The data recorded during three consecutive 2-hour periods were averaged to obtain a single mean; e.g., one mean would be based on the three 2-hour work periods beginning at 1000 hours and continuing to 1600 hours on a given day. The next mean would be based on the data recorded during the three consecutive periods that came 2 hours later, i.e., the periods beginning at 1200 hours and continuing to 1800 hours. The abscissa value against which the data were plotted was taken as the clock-time midpoint (an odd-numbered hour) of the second of the three periods involved. The data of HOPE-III were treated in a similar manner, except that the rolling means were based on four consecutive 2-hour periods instead of three.

In order to test for the significance of the differences observed among the daily levels of performance, Friedman's analysis-of-variance test for ranked data (Chi-square-r) was used. Each subject's daily mean with any specific performance measure was assigned a rank between 1 and 15 (in OPN-360 and HOPE-II), or between 1 and 30 (in HOPE-III), and these ranks were used in determining the level of significance of any over-all trends evidenced by the data of that measure. The results of these tests are given in table 5.

There were significant differences among the daily levels of the OPN-360's over-all performance of auditory vigilance, probability monitoring, and arithmetic computations, but not in their performance of warning-lights monitoring. (Because of subsequent changes in the signal characteristics, the auditory vigilance data of OPN-360 are not strictly comparable to those of HOPE-II and HOPE-III; the task was comparable in HOPE-II and HOPE-III, however, so these latter results are presented for comparison in figure 10. The data of OPN-360 are not presented.) The percentage of correct probability-monitoring signal detections decreased from the start of the study until about the eleventh day (figure 11), while the mean time required to detect the probability-monitoring signals increased from the first to the tenth or eleventh day (figure 12). The percentage of correct arithmetic computations decreased from the second day through about the fourteenth or fifteenth day (figure 15).

Although the subjects of HOPE-II were generally poorer than these of OPN-360 in performance of the auditory-vigilance task (and this was probably a function of the different signal-to-noise ratio used in the two studies), the HOPE-II subjects were generally better than those of OPN-360 in probability monitoring and arithmetic computations (see figures 11, 12, and 15). The significant differences among the daily levels of HOPE-II's over-all performance of the auditory-vigilance and probability-monitoring tasks, moreover, were essentially identical to those of the OPN-360 in being associated with declines in the levels of performance over the course of the study. In arithmetic computations, however, especially in terms of the computations made during simultaneous presentation of code-lock problems, the HOPE-II subjects demonstrated gains in performance. These gains in the percentage of correct arithmetic computations were statistically significant for problems solved with code-lock, and the gains were especially large during the early part (first 5 days) of the study (see figure 16).

The differences among the daily levels of HOPE-III's over-all performances were statistically significant in all 9 measures of performance employed (see table 5). In all cases the performance of the HOPE-III subjects improved with continued work, especially over the first 5 days of their 30-day confinement period. The final levels of performance attained with the passive tasks (auditory vigilance, probability monitoring,

Table 5

Summary of Analysis of Variance by Ranks:

Daily Levels of Individual-Task Performance

Performance Task (and Criterion Measure)	OPN-360*		HOPE-II		HOPE-III	
	Chi-square	P less than**	Chi-square	P less than**	Chi-square	P less than**
Auditory Vigilance						
(percentage correct detections)	110.69	.001	59.88	.001	59.77	.001
Probability Monitoring						
(percentage correct detections)	72.26	.001	43.98	.001	78.98	.001
(mean detection time)	82.14	.001	43.48	.001	130.69	.001
Warning-Lights Monitoring* **						
(response latency to red lights)	14.08	---	22.08	---	87.60	.001
(response latency to green lights)	19.88	---	16.72	---	133.84	.001
Arithmetic Computation						
(percentage correct: w/o code-lock)	44.64	.001	22.05	---	131.45	.001
(percentage correct: with code-lock)	n.a.	n.a.	47.44	.001	167.49	.001
Target Identification, Individual						
(percentage correct: w/o code-lock)	n.a.	n.a.	12.28	---	84.13	.001
(percentage correct: with code-lock)	n.a.	n.a.	14.73	---	85.19	.001

* Adopted from Adams and Chiles (ref. 3)

** Based on df of 14 for OPN-360 and HOPE-II, and on df of 29 for HOPE-III.

*** A normalized scale transformation was used with these latency data (see Alluisi, Hall, and Chiles, ref. 5, p. 25).

n.a. — not applicable.

and warning-lights monitoring) was highest in the case of the HOPE-III subjects relative to those of the other two groups; however, the subjects in HOPE-II maintained the highest levels of performance in arithmetic computations. Because of the differences in the levels of visual noise employed, the individual target-identification data of HOPE-II and HOPE-III are not strictly comparable, but performance in target identifications with simultaneous presentation of code-lock problems gradually approached that obtained without code-lock probelms as both groups gained additional practice.

An indication of the consistency of the trends for individual subjects is given in the data of table 6 where there are listed the rank-order coefficients of correlation (rho's) between the individual subjects's means of daily performance measures and the days of the study. The time measures for the probability-monitoring and warning-lights tasks were converted to speed measures so that the interpretation of positive and negative rho's would be consistent over tasks. Thus, a positive correlation represents an improvement with continuation of the study; a negative correlation represents decrements in performance.

Of the 66 coefficients (rho's) reported for OPN-360, 50 are significantly different from zero; of these 50 significant rho's, 45 are negative and indicate that the associated performances decreased with time. All 5 of the significant, positive rho's occurred among 3 of the 6 subjects in Crew B, and 12 of the 16 non-significant rho's occurred among the members of the same crew.

Twenty-five of the 54 coefficients reported for HOPE-II were statistically significant, and of these 25, 18 indicated decrements in performance and 7 indicated improvements. There were no significant trends evidenced in the data of the individual responses to the target identification task; this is not surprising, inasmuch as different levels of visual noise were used during the three 5-day blocks of experimentation and the individual subject's level of noise was changed from day to day. Of the 30 coefficients for the passive tasks (auditory vigilance, probability monitoring, and warning-lights monitoring) 17 negative correlations were significant; there were no significant positive correlations. Thus, performance appears to have declined generally on these passive tasks. However, of the 12 coefficients for the active arithmetic-computation task (6 with and 6 without the code-lock task) only one significant rho was negative; the remaining 7 significant rho's were positive and represent improvements in performance. Two of the 6 subjects (nos. 3 and 5) showed significant decrements only in auditory vigilance, and either no changes or improvements on the other tasks.

There are 90 rho's for HOPE-III based on the performance of 10 crewmembers on each of the 9 individual-performance measures. Of these 90, 63 are statistically significant; of these 63 significant coefficients, only 5 are negative. Three of these negative correlations occurred in auditory vigilance. The remaining two occurred with one subject on both conditions of the target-identification task; that subject was also one of the three who showed a decrement in auditory vigilance.

Within-Day Trends - The evaluation of the presence and nature of diurnal variations in the individual performance tasks was based on data combined over 5-day blocks of the experiment. To take HOPE-II as an example, the performance data for corresponding 2-hour periods of the day were combined for days 1 to 5 to yield a mean daily performance curve for each measure for the first five days of the study. Mean daily curves for days 6 to 10, and days 11 to 15, were derived in the same manner. The within-day data of OPN-360 and HOPE-III were handled the same way, except that there were six 5-day blocks in HOPE-III.

Friedman's test was used to test the statistical significance of the observed differences among the within-day means for each of the mean 5-day curves and for the entire

Table 6

Rank-Order Coefficient of Correlation (Rho) of Each Subject's Daily Performance with Days of the Studies

Group	Subject	Aud. Vigil. (% correct)	Prob. Monit. (% correct)	Prob. Monit. (speed)	Warning Lights Red (speed)	Warning Lights Green (speed)	Arith. Comp. w/o Code Lock	Arith. Comp. with Code Lock	Target ID w/o Code Lock	Target ID w/ Code Lock
		Performance Task and Measure								
OPN-360 *	1A	-.96	-.98	-.96	-.73	-.93	-.84	n.a.	n.a.	n.a.
	2A	-.84	-.90	-.87	-.94	-.93	-.94	n.a.	n.a.	n.a.
	3A	-.94	-.59	-.78	(-.11)	(-.30)	-.52	n.a.	n.a.	n.a.
	4A	-.50	-.78	-.85	(-.34)	(-.25)	-.48	n.a.	n.a.	n.a.
	5A	-.88	-.90	-.88	-.86	-.78	-.96	n.a.	n.a.	n.a.
	1B	-.96	-.84	-.93	-.68	(+.09)	(+.38)	n.a.	n.a.	n.a.
	2B	-.92	-.52	-.55	+.64	+.53	-.87	n.a.	n.a.	n.a.
	3B	-.91	(-.28)	-.50	-.59	-.57	(-.04)	n.a.	n.a.	n.a.
	4B	-.63	(-.19)	(-.30)	+.78	+.50	-.83	n.a.	n.a.	n.a.
	5B	-.50	(-.04)	(+.06)	(+.21)	(+.08)	(+.10)	n.a.	n.a.	n.a.
	6B	-.96	-.54	-.75	+.60	(+.34)	-.51	n.a.	n.a.	n.a.
HOPE-II *	1	-.95	-.86	-.83	(-.19)	-.70	(+.03)	+.68	(-.05)	(+.09)
	2	-.85	-.66	-.70	-.49	(+.31)	+.50	+.70	(-.31)	(+.08)
	3	-.74	(-.07)	(+.11)	(+.25)	(+.38)	(-.02)	+.58	(-.01)	(+.37)
	4	-.63	-.85	-.74	-.56	(+.30)	-.46	(-.15)	(-.24)	(+.02)
	5	-.49	(+.20)	(+.20)	(-.31)	(-.03)	+.59	+.71	(-.16)	(+.18)
	6	-.88	-.74	-.79	(-.24)	(-.08)	(+.23)	+.62	(+.02)	(+.02)
HOPE-III **	1A	+.32	(+.25)	+.39	+.57	+.81	+.74	+.83	(+.23)	+.48
	3A	+.38	+.51	+.43	(+.20)	+.48	+.66	+.70	(-.02)	(+.26)
	5A	+.41	+.47	+.62	+.39	+.88	+.74	+.77	(+.22)	+.33
	7A	+.48	+.58	+.94	+.58	+.77	+.48	+.48	(+.18)	+.54
	9A	+.72	+.51	+.70	+.55	+.72	+.52	+.68	(+.23)	(+.28)
	2B	-.83	(+.29)	+.69	+.55	+.74	+.82	+.76	(+.09)	(-.01)
	4B	(+.24)	+.56	+.89	+.41	+.79	+.78	+.85	(+.14)	+.31
	6B	-.48	+.56	+.87	+.46	(+.28)	+.64	+.90	(+.03)	(+.22)
	8B	(+.10)	(+.29)	(+.24)	(+.02)	+.37	+.42	+.53	(+.01)	(+.04)
	10B	-.59	+.38	(+.07)	(-.02)	+.44	(+.07)	(+.22)	-.40	-.53

N.B. Parentheses are used to enclose non-significant values of rho.

* $\underline{N}$ = 15 in each case, so for a one-tailed test, rho = .44 is associated with $\underline{P}$ = .05, and rho = .62 is associated with $\underline{P}$ = .01.

** $\underline{N}$ = 30, so for a one-tailed test, rho = .31 is associated with $\underline{P}$ = .05, and rho = .43 is associated with $\underline{P}$ = .01.

period of experimentation. For the purpose of these analyses, which are summarized in table 7, a mean for each task was computed for each subject and each 2-hour period during which that subject was on duty. Thus, there were 8 such duty periods per day for the OPN-360 and HOPE-II subjects, and 6 periods per day for the HOPE-III subjects. To take HOPE-II as an example, the 8 duty periods were ranked from 1 to 8 for each subject, and these ranks were then used in the computation of the Friedman statistic. However, because of the lack of meaningfulness of plots of the <u>mean ranks</u> thus obtained, the curves showing the within-day performance changes are based on rolling means of the <u>raw data</u>. Unfortunately, the smoothing effect of the rolling-mean technique tends in some instances to obscure the significant trends revealed by the Friedman analysis (see figure 24, days 1-5 for HOPE-II).

Attending first to the data of OPN-360, a significant within-day effect is seen to be present in all individual-performance measures when the scores are averaged over the entire 15-day period, and also in 10 of the 18 separate 5-day mean curves. With the same 6 measures, on the other hand, only 1 significant within-day effect is present in HOPE-III's data when averaged over the 15-day period (arithmetic computations without code lock); of the 18 separate 5-day mean curves of tasks that are common to HOPE-II and OPN-360, 2 demonstrated significant within-day effects. Four of the 18 corresponding curves are significant in the data of HOPE-III. Thus, a simple count of the number of significant trends indicates that within-day changes in the individual-performance measures were greatest with OPN-360, least with HOPE-II, and intermediate with HOPE-III. The relative ordering of the latter two crews is attested to further by the fact that when the HOPE-II data are compared with the first 15 days of the HOPE-III study, of the 27 separate mean daily curves (3 blocks x 9 measures), 4 showed significant variation for HOPE-II, whereas 8 were significant for HOPE-III.

When scores were averaged over the entire period of experimentation, 3 of the 9 measures showed significant within-day changes with the 15-day confinement of HOPE-II, and 7 of the 9 showed significant within-day changes with the 30-day confinement of HOPE-III. Where the variations in within-day performance were significant, shifts in the times of maximum performance from one 5-day block to another were somewhat apparent; these shifts, like those identified in the psychophysiological measures, appear to have been much more consistent with the crews on the 4-2 schedule (especially with the OPN-360 crews) than with the 4-4 schedule (HOPE-III). For example, in the percentage of correct arithmetic computations without simultaneous presentation of code-lock problems (figure 24), the period of best performance for OPN-360 shifts from the late evening hours to the early morning hours, whereas this is not so clearly the case with either HOPE-II or HOPE-III.

PERFORMANCE MEASURES, GROUP TASKS

As indicated earlier, two of the tasks in the battery were crew or group-performance tasks. One of these two tasks, code-lock solving, can be scored only as a group task, whereas the other, target identification, can be scored in terms of both individual and group-dependent performances. The individual-performance aspects of the target-identification task have been reported in the preceding section. The group-performance aspects, the "commander's final decision," will be reported in this section along with the code-lock task.

<u>Target Identification: Commander's Final Decision</u> - The mean percentages of correct final decisions made by the HOPE-II and HOPE-III crew commanders in target identification are presented in figures 26, 27 and 28. As was the case with individually scored target identifications, between-day trends over the 15-day or 30-day periods are not too meaningful because of the 3 different visual-noise conditions employed during

Table 7

Summary of Analysis of Variance by Ranks: Within-Day Levels of Individual-Task Performance

Task (and Measure)	Day Blocks	OPN-360*		HOPE-II		HOPE-III	
		Chi-square	P less than**	Chi-square	P less than**	Chi-square	P less than**
Auditory Vigilance (% correct)	1-5	11.30	---	2.44	---	4.46	---
	6-10	14.01	---	8.39	---	9.50	---
	11-15	12.25	---	18.85#	.01	2.00	---
	16-20	n.a.	n.a.	n.a.	n.a.	5.47	---
	21-25	n.a.	n.a.	n.a.	n.a.	3.83	---
	26-30	n.a.	n.a.	n.a.	n.a.	9.59	---
	1-15 or 1-30	16.03	.025	6.02#	---	15.94	.01
Probability Monitoring (% correct)	1-5	13.86	---	7.03	---	14.97	.02
	6-10	15.52	.05	12.86	---	6.49	---
	11-15	13.81	---	11.92	---	4.19	---
	16-20	n.a.	n.a.	n.a.	n.a.	0.20	---
	21-25	n.a.	n.a.	n.a.	n.a.	8.95	---
	26-30	n.a.	n.a.	n.a.	n.a.	1.38	---
	1-15 or 1-30	16.04	.025	10.56	---	4.29	---
Probability Monitoring (Time)	1-5	17.27	.02	9.67	---	8.39	---
	6-10	14.21	.05	6.72	---	13.65	.02
	11-15	19.24	.01	4.11	---	18.96	.005
	16-20	n.a.	n.a.	n.a.	n.a.	5.70	---
	21-25	n.a.	n.a.	n.a.	n.a.	16.27	.01
	26-30	n.a.	n.a.	n.a.	n.a.	7.30	---
	1-15 or 1-30	23.61	.005	5.22	---	23.25	.001
Warning-Lights Monitoring (Red Latency)	1-5	30.20	.001	7.98	---	3.69	---
	6-10	16.00	.05	3.72	---	5.07	---
	11-15	23.55	.005	1.43	---	19.24	.005
	16-20	n.a.	n.a.	n.a.	n.a.	14.86	.02
	21-25	n.a.	n.a.	n.a.	n.a.	11.54	.05
	26-30	n.a.	n.a.	n.a.	n.a.	3.90	---
	1-15 or 1-30	22.76	.005	6.95	---	21.57	.001
Warning-Lights Monitoring (Green Latency)	1-5	12.73	---	5.38	---	3.90	---
	6-10	20.27	.01	1.78	---	5.33	---
	11-15	12.67	---	12.56	---	10.00	---
	16-20	n.a.	n.a.	n.a.	n.a.	8.71	---
	21-25	n.a.	n.a.	n.a.	n.a.	15.66	.01
	26-30	n.a.	n.a.	n.a.	n.a.	9.49	---
	1-15 or 1-30	21.55	.005	9.67	---	10.76	---

Arithmetic Computations (% correct: w/a Code-Lock)	1-5	21.20	.005	14.33	.05	9.23	---
	6-10	9.83	---	9.36	---	3.25	---
	11-15	19.93	.01	8.24	---	6.86	---
	16-20	n.a.	n.a.	n.a.	n.a.	7.40	---
	21-25	n.a.	n.a.	n.a.	n.a.	10.93	---
	26-30	n.a.	n.a.	n.a.	n.a.	10.26	---
	1-15 or 1-30	24.41	.001	16.33	.025	16.67	.01
Arithmet ic Computatians (% correct: with Code-Lack)	1-5	n.a.	n.a.	21.40	.005	21.87	.001
	6-10	n.a.	n.a.	10.00	---	17.49	.005
	11-15	n.a.	n.a.	3.88	---	9.49	---
	16-20	n.a.	n.a.	n.a.	n.a.	17.27	.005
	21-25	n.a.	n.a.	n.a.	n.a.	1.53	---
	26-30	n.a.	n.a.	n.a.	n.a.	5.08	---
	1-15 or 1-30	n.a.	n.a.	16.18	.025	21.23	.001
Target ID, Individual (% correct: w/o Code-Lock)	1-5	n.a.	n.a.	10.26	---	7.66	---
	6-10	n.a.	n.a.	5.14	---	12.67	.05
	11-15	n.a.	n.a.	8.65	---	10.67	---
	16-20	n.a.	n.a.	n.a.	n.a.	16.17	.01
	21-25	n.a.	n.a.	n.a.	n.a.	14.89	.02
	26-30	n.a.	n.a.	n.a.	n.a.	12.00	.05
	1-15 ar 1-30	n.a.	n.a.	5.54	---	14.50	.02
Target ID, Individual (% correct: with Code-Lock)	1-5	n.a.	n.a.	16.50	.025	5.89	---
	6-10	n.a.	n.a.	3.65	---	5.72	---
	11-15	n.a.	n.a.	5.74	---	17.85	.005
	16-20	n.a.	n.a.	n.a.	n.a.	7.97	---
	21-25	n.a.	n.a.	n.a.	n.a.	7.09	---
	26-30	n.a.	n.a.	n.a.	n.a.	11.20	.05
	1-15 ar 1-30	n.a.	n.a.	19.18	.01	28.70	.001

* Adopted fram Adams and Chiles (ref. 3).

** Based on df af 7 for OPN-360 and HOPE-II, and on df of 5 far HOPE-III.

\# Data of 5 subjects anly; the sixth subject develaped an ear infection during this period and as a result lost auditory acuity.

n.a. Not applicable.

the three 5-day blocks of experimentation. (The noise levels listed in figures 17, 18, 19, 26, 27, and 28 are weighted arithmetic means of the percentages associated with the high and the low noise levels for the specified 5-day blocks.)

Two things are immediately apparent in comparing individual performance with group performance (the commander's group decision) with respect to the average number of correct responses for both HOPE-II and HOPE-III (figure 17 versus figures 26 for HOPE-II, and figures 18 and 19 versus 27 and 28 for HOPE-III). First, the level of performance achieved by the commander in making the group decision generally exceeded the average level for the individual responses. Second, the differences in the performance of this task with code-lock as compared to its performance without code-lock is smaller for group performance than for individual performance.

Perhaps of greater interest are the decision methods apparently used by the commander in both studies as inferred by the experimenters from their observations of the patterns of responding displayed at the experimenter's station. All the commanders seemed to behave in the same general way. When the commander was one of the two crewmembers assigned the lower level of visual noise on a given day, he would typically enter his individual judgment as the group decision without regard to the responses of the other crewmembers and often made this final entry before all of the others had responded. Occasionally, when certain of the crewmembers was assigned the lower noise level, the commander would wait until that crewmember responded before entering his final decision. If their individual responses were not in agreement, the commander would often call over the intercom to the crewmember in question requesting a "level of confidence." If the reply indicated a level of confidence that approached certainty, the commander might enter that crewmember's answer as the group decision (depending, apparently, on his own level of confidence with regard to his individual judgment).

When the commander was one of the crewmembers assigned the higher noise level, he generally entered as the group decision the judgment of the two men assigned the lower noise if they agreed; the use of this approach by the commander was especially noticeable when the two noise levels differed greatly. When the noise levels did not differ much, the commander would be likely to enter his own individual judgment as the group decision, especially if his individual response agreed with that of certain other crewmembers.

Thus, when the high level of noise was relatively more severe than the low level, the individual responses of the crewmembers assigned the higher level tended to be generally ignored by the crew commander. Likewise, the decisions of some of the crewmembers were ignored nearly all of the time, even when they were assigned the lower noise level. Since the individual crewmembers were informed as to the commander's group response through cue lights, a given subject would know that his response was being completely ignored when the commander entered a final response before that specific subject had responded and, of course, he would always know when the commander disagreed with his response.

Code-Lock Solving - Mean time per response, mean percentage of erroneous responses (i.e., responses made out of sequence), and the mean number of code-lock problems solved per minute were the three principal criteria used to evaluate performance on the code-lock task. Each of these criteria was scored separately for performance of the task with simultaneous presentations (a) of arithmetic-computation problems, (b) of target-identification problems, and (c) of the monitoring and vigilance tasks only.

Also, since the subjects were required to re-enter their solutions following a 30-second delay after solving the problem, all criterion measures were analyzed separately for the "first solutions," the "second solutions," and the "composite" or total of both solutions combined. The short-term memory required to make rapid and errorless second solutions was expected to be especially sensitive to diurnal variations, and it was feared that use of the composite score without verification of its sensitivity might lead to loss of some sensitivity. However, separate analyses of variance carried out on each of the three scoring techniques yielded essentially identical results for each of the main effects and the interactions. Accordingly, only the data of the composite scores are presented here, these having been selected for use over either the first-solution or the second-solution data because they alone include all the code-lock data.

A word of explanation might be in order concerning the use of the three criteria selected: mean time per response, mean percentage of erroneous responses, and mean number of code-lock problems solved per minute. In an earlier study (Alluisi, Hall, and Chiles, ref. 5), the results obtained in code-lock solving with 5 information measures of performance were factor analyzed. The analysis (see ref. 5, table 16, page 37) indicated that two measures could be used to represent all the informational data collected with the code-lock task. These two measures were identified as the relative information rate while responding, R(respdg), and the equivocation rate while responding, E(respdg). Accordingly, both measures were computed from the data of HOPE-II, but they appeared, on inspection, to be quite highly correlated with the two criteria already selected for use -- the more easily interpreted criteria of latency (mean time per response) and errors (mean percentage of erroneous responses).

In order to verify this observation, and to obtain a better understanding of the relations among the measures of code-lock solving, another factor analysis was computed. Intercorrelations were computed among the 5 measures of (a) mean time per response, (b) mean percentage of erroneous responses, (c) R(respdg), (d) E(respdg), and (e) the mean number of code-lock problems solved per minute -- this last measure being also a linear transformation of the information measure of the relative information rate per period, R(period), as defined previously (see Alluisi, Hall, and Chiles, ref. 5, pp. 29-30). In addition to its being fairly easy to interpret (with "instantaneous" responses the subjects could at best solve only 2 code-lock problems per minute because of the built-in 30-second delay between successive problems), this last measure has the advantage of being essentially identical in its previous factor loadings to the two remaining information measures excluded from the current analysis (rate of information transmission per period, and rate of information transmission while responding; cf. Alluisi, Hall, and Chiles, ref. 5, page 37). For these reasons it was included in the current analysis.

The intercorrelations were computed from the data of HOPE-II; 45 points were used for each correlation, these being the data obtained on each of the 15 days under each of the three conditions of responding. The intercorrelations, rotated factor loadings, residuals, and communalities resulting from this analysis are presented in table 8. It is apparent that there are two multiple-group factors, I and II. Factor I will be identified as speed; it has a high negative loading associated with the mean time per response, or latency, and a high positive loading associated with the equivocation rate while responding. Factor II will be identified as accuracy; it has a high negative loading associated with the mean percentage of erroneous responses, and a high positive loading associated with the relative information rate while responding. The remaining fifth measure of performance included in the analysis, the mean number of code-lock problems solved per minute (or the relative information rate per period), is equally loaded on Factors I and II. That is to say, it is equally loaded on the speed and accuracy factors and constitutes, therefore, a composite measure of code-lock performance -- a measure that combines speed and accuracy, and that does the combining equally well for these two pertinent aspects of performance.

Table 8

Intercorrelations, Rotated Factor Loadings, Residuals, and Communalities of Five Measures of Performance in Code-Lock Solving*

Measure	Code Number	Measure (in Code Number)					Rotated Factor Loading	
		1	2	3	4	5	I	II
Mean Time per Code-Lock Response	1	(970)	-005	-052	-010	-041	-90	-40
Equivocation Rate While Responding	2	-980	(992)	050	001	042	95	30
Mean Number of Code-Lock Problems Solved per Minute	3	-942	910	(912)	-010	053	70	65
Relative Information Rate While Responding	4	-651	571	818	(992)	001	30	95
Mean Percentage of Erroneous Code-Lock Responses	5	544	-466	-707	-890	(872)	-25	-90

* Figures in lower-left half of matrix are intercorrelations; those in the upper-right half are residuals. Communalities are given in parentheses on the diagonal; these are consistent with the reliabilities reported previously (see Alluisi, Hall, and Chiles, ref. 5, table 3, p. 7). For ease of reading, the decimal point that should precede each entry in the table has been omitted.

The mean percentages of erroneous responses under the three different response conditions (i.e., with simultaneous presentation of arithmetic computations, of target identification, and of the monitoring and vigilance tasks only) are presented in figures 29 and 30 for HOPE-II and HOPE-III, respectively. The comparable data for mean time per code-lock response under the same three response conditions are presented in figures 31 and 32, and the data for mean number of code-lock problems solved per minute are given in figures 33 and 34. Summaries of analyses of variance of these data are given in tables 9 and 10 for HOPE-II and HOPE-III, respectively.

In the case of HOPE-II, 3 primary variables were examined in each analysis: response conditions, days of experimentation, and the 2-hour performance periods within a day. (It should be noted, incidentally, that differences in time/response among the response conditions are presumably a function of differential time-sharing requirements imposed by the simultaneously performed tasks; differences among the response conditions with the other two criteria would presumably arise because of these time-sharing requirements.) It is apparent from the data of table 9 that all three of the major variables produced significant differences in performance of code-lock solving with all three criteria of performance. In addition, all first-order interactions were significant in the cases of mean time/response and mean numbers of problems solved/minute, but only the interaction of response conditions with days of experimentation was significant with the error criterion.

Table 9

Summary of Analysis of Variance of Composite Scores of HOPE-II Code-Lock Responses

Source of Variation	df	Percentage Errors		Time/Response		Problems Solved/Minute	
		Mean Square	F	Mean Square	F	Mean Square	F
Response Conditions (C)	2	916.8	105.38***	2584.75	2482.95***	6.3915	702.36***
Days of Responding (D)	14	20.9	2.40**	82.98	79.71***	0.1025	11.26***
2-Hour Periods within a Day (H)	11	72.7	8.36***	49.10	47.16***	0.0934	10.26***
C x D	28	14.5	1.67*	59.52	57.17***	0.0259	2.85***
C x H	22	12.5	1.44	11.18	10.74***	0.0241	2.65***
D x H	154	10.5	1.21	7.09	6.81***	0.0129	1.42**
C x D x H	308	8.7	---	1.04	---	0.0091	---
Total	539	---	---	---	---	---	---

* P less than .025

** P less than .005

*** P less than .001

Since two different crews were used in HOPE-III, teams of subjects constituted a dimension of random variation in the analysis of the HOPE-III data. Also, the alternation of crews on duty according to the 4-4 schedule used in HOPE-III precluded the making of any straight-forward interpretation of an hours-within-days effect, and, therefore, this effect was not included in the analysis. The analysis of the HOPE-III data (table 10) indicated that with respect to the latency measure and the composite (number of problems solved/minute), significant differences were obtained for both major dimensions and all interactions. With the erroneous-response criterion, on the other hand, only the differences in performance among the response conditions were significant, along with the interaction of subject teams with the days of experimentation.

The appropriate figures, in conjunction with the data of tables 9 and 10, reveal the following general results on the code-lock task: (a) The highest mean percentages of erroneous responses were made under the response condition involving simultaneous presentation of target identification; the lowest were made under the condition involving simultaneous presentation of monitoring and vigilance tasks alone (figures 29 and 30). (b) These differences were greater with HOPE-II than with HOPE-III, as were also the day-to-day fluctuations in errors. (c) As confinement and practice continued, the performance of the HOPE-III crews improved to a significant extent with the time criterion (figure 32), but not so greatly, or perhaps not at all, with the error criterion (figures 30); the composite score of problems solved/minute (figure 34) indicates HOPE-III continued to improve on code-lock solving with simultaneous presentation of the two active tasks (arithmetic computations and target identification), but not (after the tenth day) with concurrently presented monitoring and vigilance alone. (d) There was a significant increase in the number of errors over days for HOPE-II, as well as a significant slowing of the rate of responding; this is perhaps best illustrated in the composite speed-and-accuracy scoring provided by the criterion of number of problems solved/minute (figure 33), where general downward trends are evidenced for HOPE-II under all three response conditions. (e) The subjects of HOPE-II showed significant variations in their within-day performance with all three criterion measures; (but these variations were small in magnitude). The changes in within-day performance for the HOPE-III crews were not subjected to a statistical analysis, but they appear to be relatively minor.

DISCUSSION

The evaluation of the results of HOPE-II and HOPE-III, as well as their relation to the earlier 15-day study, Operation-360 (OPN-360), should take account of several important differences in the studies other than the major parameters that were deliberately manipulated. For example, there were three differences between the first and second crews tested in the OPN-360: (a) The first crew contained only 5 subjects, whereas the second contained 6. (b) The second crew had approximately 1 hour additional per day available for sleep because they were permitted to eat their meals during the low-performance portion of their duty period, whereas the first crew had to eat all their meals during off-duty time. The third, and perhaps most important difference between the two crews of OPN-360, was that (c) the second crew had volunteered for the study, whereas the first crew had been the only group "available to volunteer."

There were likewise three major differences between the OPN-360 and the HOPE-II and HOPE-III studies. First, in the latter two studies, the subjects were required to perform two group interaction tasks, whereas such tasks had not been represented in the earlier OPN-360 studies. Second, each subject volunteered as an individual to serve in the later studies, presumably without the kind of social-pressure phenomena that might have existed in the regularly constituted groups that were being asked to volunteer as units to serve in OPN-360. Third, and finally, the HOPE-II and HOPE-III subjects were asked to attempt to minimize the previously demonstrated diurnal performance effects.

Table 10

Summary of Analysis of Variance of Composite Scores of HOPE-III Code-Lock Responses

Source of Variation	df	Percentage Errors		Time/Response		Problems Solved/Minute	
		Mean Square	F	Mean Square	F	Mean Square	F
Response Conditions (C)	2	112.19	701.19**	8201.95	46.21*	10.2681	61.23*
Days of Responding (D)	29	9.67	1.30	219.36	10.26***	0.2265	17.83***
Teams of Subjects (T)	1	9.92	n.t.	967.34	n.t.	0.6801	n.t.
C x D	58	3.91	---	39.37	2.33***	0.0256	1.94*
C x T	2	0.16	---	177.49	37.42***	0.1677	30.49***
D x T	29	7.45	2.05**	21.38	4.39***	0.0127	2.31***
C x D x T	58	4.19	1.15	16.91	3.47***	0.0132	2.40***
Residual (including all periods nested in groups)	900	3.63	---	4.87	---	0.0055	---
Total	1079	---	---	---	---	---	---

* P less than .C25

** P less than .005

*** P less than .001

n.t. - no test possible

The subjects of all three studies differed on two additional points. First, the age ranges differed: the ages ranged from 26 to 43 years for both crews in OPN-360; they ranged from 19 to 22 years in HOPE-II, and from 23 to 26 years in HOPE-III. The second difference has been inferred and relates to the motivational continuum. The subjects in all three experiments are assumed to have regarded the studies equally with respect to their value to the Air Force, and therefore, this source of motivation is assumed to be about equal over all groups. Because of the age differentials, however, the subjects of HOPE-II and HOPE-III can be assumed to have regarded the experiments as having a greater potential impact on their Air Force careers than the subjects of OPN-360. Also, the group-performance tasks presented during HOPE-II and HOPE-III can be assumed to have provided some social facilitation to performance, i.e., these tasks probably have an incremental effect on motivation by the stimulation of interest as well as the stimulation of possible group pressures "to stay with it."

For all these reasons, and especially because they had volunteered as individuals, the HOPE-II and HOPE-III groups were assumed to have additional sources of motivation beyond those of OPN-360. Finally, the subjects of HOPE-II (who came from the Air Force Academy) exhibited an esprit de corps and a desire to succeed on behalf of the Academy that was not likely to be present for the subjects in the other experiments. Thus, it is generally likely that the HOPE-II subjects had the highest level of motivation, the HOPE-III subjects had the next highest level, and subjects of OPN-360 had the next highest level. However, it should be emphasized that this is strictly a relative question and should not be construed as a suggestion that the motivation of the subjects in OPN-360 was poor. There were many observational indications that the motivational element common to all groups, that the study "is of value to the Air Force," is a potent motivational factor among dedicated military subjects such as those who served in these studies.

Another important difference between the HOPE-II crew and the crews of subjects used in the other two studies may account partially for the differences in the trends of the psychophysiological measures across the days of the experiments. Considering the specific trends, the crews in OPN-360 apparently started off at a relatively high level of psychophysiological activation (primarily as evidenced by their high pulse rates) and then the level gradually lowered through the tenth day. As reflected in table 3, this was less pronounced for the second crew in OPN-360 than for the first, presumably because the second crew knew that a group of subjects had already successfully completed what they were being asked to do.

In contrast, the HOPE-II subjects started out at a relatively low level of activation which gradually increased over the course of the confinement period. The interpretation of this over-all (though not significant) increase in pulse rate with the HOPE-II crew must take account of the known adaptation effects that result when individuals acclimatized to a higher altitude move to a lower altitude. Specifically, immediate lowering of pulse rate should be expected to occur when Academy cadets are shifted from the Air Force Academy (about 7000 feet Mean Sea Level) to the higher partial pressure of oxygen at Marietta, Georgia (about 1000 feet Mean Sea Level) as they were in HOPE-II, (ref. 7).

Thus, it may well be that this increase in pulse rate over the first few days was a function of homeostatic adjustment to the atmospheric conditions rather than a reflection of an increasing level of activation. This would appear to be the most reasonable explanation for the difference between the psychophysiological trends exhibited by the crews in OPN-360 (who came from an altitude of about 1000 feet) and the HOPE-II crew. The generally decreasing pulse rates of the HOPE-III subjects indirectly add weight to this notion; those subjects came from altitudes more like that to which the subjects in OPN-360 were accustomed (about 3000 feet for the Able crew, and 1000 feet for the Baker crew).

The subjects in all three studies showed similar, very marked diurnal variations in the psychophysiological measures. The stability of the measures over studies, as well as within studies, suggests that the measures are independent of (a) the differences in the work-rest schedules used, (b) the presence or absence of the group-performance tasks, (c) specific knowledge of and instructions about the presumed existence of such diurnal variations, (d) whether the measures are obtained by the experimenter or by the subject himself, and (e) the differences in the subject populations sampled. However, there is evidence in the HOPE-III data that when the period of confinement is extended beyond 25 days the magnitude of these within-day fluctuations decreases. Specifically, the within-day variations for the final 5 days of the HOPE-III study were not significant, whereas those for the immediately preceding five days were. This slow rate of adaptation is expected from previous data concerning the psychophysiological effects of atypical sleep-wakefulness schedules (ref. 6, pp 12-15 and 20-22). Unfortunately, we do not have data directly relevant to whether or not the observed flattening of the within-day curves is dependent on the specific (4-4) schedule used.

The 4-hour shift in the period of daily maximum activation shown by the subjects in OPN-360 was also found in HOPE-II. However, as noted earlier, the HOPE-III subjects did not show this magnitude of shift over the first 15 days. Specifically, in the case of pulse rate they showed at most a 2-hour shift over days 1-15 and then returned to their original peak hour by the final 5 days (days 26-30). In the case of axillary temperature, the HOPE-III subjects shifted 2 hours during days 1-15 and then another 2 hours during days 16-30. The only readily available explanation (other than chance) for the lack of consistency between pulse rate and axillary temperature for HOPE-III, would be some idiosyncratic characteristic of the 4-4 schedule as compared to the 4-2 schedule. One could argue, though the supporting evidence is not available, that the shift is primarily a product of some progressive or accumulative effect of the mild degree of sleep deprivation produced by the 4-2 schedule. The contribution of such an argument is perhaps limited to its heuristic value. The possible implications for psychophysiological theory of considering the lack of consistency between the pulse rate and axillary temperature for HOPE-III to be other than chance are rather profound.

Turning to the performance measures, a rather unclear picture emerges in trying to relate the between-days effects for performance and the between-days effects for the psychophysiological measures. The subjects in OPN-360 showed a declining activation level and, with the exception of two subjects, also showed declining levels of performance. The HOPE-III subjects showed the declining activation level, but with the exception of the auditory-vigilance task and one subject's performance on two other tasks, they showed an increasing level of performance over the duration of the study. The HOPE-II subjects showed an increasing activation level, but a decreasing level of performance on auditory vigilance, probability monitoring, and code-lock solving. As noted earlier, it is not clear that the changes in the psychophysiological measures of HOPE-II should be interpreted as changes in activation level; even if this is the correct interpretation, however, it could still be that the impact of the 4-2 schedule is so great, the subjects are unable to hold their own on these three tasks in spite of the increased activation level.

In the OPN-360, the subjects had shown well defined diurnal variations on most of the performance tasks; these variations appeared to keep pace closely with the psychophysiological variations. This was the reason that special instructions were given to both HOPE-II and HOPE-III to attempt to eliminate such performance variations. The HOPE-II subjects were amazingly successful in preventing such diurnal cycling. Of the tasks that were comparable to those performed in OPN-360, only the arithmetic task (during the first 5 days and averaged for the 15 days) showed significant within-day variations; furthermore, the magnitude of these effects (though significant at the .05 and .025 levels, respectively) were very minor as compared with the data of OPN-360.

The instructions for HOPE-III, as compared with those for HOPE-II, did not place quite as much emphasis on the avoidance of cycling. This may partially explain why the HOPE-III data showed significant diurnal performance effects during one or more five-day periods on all tasks, and significance on all except the probability-monitoring (percentage correct) and green warning-lights tasks for the 30-day averages. However, it is important to note that with the exception of the arithmetic task the performance of the HOPE-III subjects equaled or was superior to that of the HOPE-II subjects at all times. Thus, in terms of system readiness, the HOPE-III system was at a higher level of preparedness than was the HOPE-II system, generally, even though the HOPE-III system showed somewhat greater variability.

Another point of importance in considering the presence of diurnal cycling in performance is the relation of such cycles to the effective workload placed on the operator. The portion of the task program that places the greatest demands on the subject is probably the 15-minute period during which the arithmetic and code-lock tasks are performed simultaneously along with the monitoring and vigilance tasks. The subjects were presented a new arithmetic problem every 20 seconds; except for the 30-second delays between the first and second solutions and between successive code-lock problems, the subjects would have to respond continuously on the code-lock task if they were to minimize the solution time.

However, the natures of these two tasks precluded their being performed simultaneously in the literal sense. Rather, in order to optimize performance during this high load period, the subjects had to complete their arithmetic problems as rapidly as possible and then they had to spend the remainder of the 20-second period working on the current code-lock problem. During the first few days of the HOPE-II and HOPE-III studies, the subjects had to brush up on their elementary arithmetic in order to minimize the time required to complete each individual problem. The subjects who tended to be slower on the arithmetic task were under particular pressure to speed up, since the code-lock problems could not be pursued effectively unless all subjects were available to respond. Because of this it was apparent that the highest effective workload placed on the subjects came during these early days of the studies when arithmetic and code-lock were presented together. Inspection of figures 16 and 25 suggests that performance during these periods of high effective workload was sensitive to the diurnal cycling of alertness. This inference is supported further by the fact that the arithmetic task, when performed without simultaneous code-lock solving, did not show diurnal cycling of performance to any appreciable degree (see figures 15 and 24).

The impact of these first few days was compounded by the fact that this was a period of adjustment to the unusual work/rest schedule -- a period that typically results in at least some loss of sleep. Taken together, these considerations suggest that the effects of performance demands which approach the overload point will depend upon the location of that period on the diurnal-performance curve. At this time, the parameters of this curve can probably be estimated best from the curves of the psychophysiological measures, since, where performance cycling does occur, it appears to parallel rather closely the psychophysiological cycling.

The major differences among the studies, in terms of the impact on the individual subjects, are revealed in the individuals' between-day trends in performance. Three of the five subjects in the first OPN-360 crew showed decrements on all tasks, and the other two subjects showed decrements on all but the two warning-lights tasks (actually, they showed decrements here too, but the correlations were not significant). Four of the six subjects in the second OPN-360 crew showed decrements on most tasks; of the remaining two subjects, one showed decrements only on auditory vigilance and arithmetic, and one on auditory vigilance only. The difference between the results for these two crews was presumably produced primarily by the fact that the second crew had approximately 1 hour additionally available for sleep each day; a possible second reason for their better performance may have derived from their presumably different reactions to the situation because (a) they were volunteers and (b) they knew of the previous, successful completion of the study by the first crew.

Two of the six subjects in HOPE-II showed significant decrements on only one task (auditory vigilance); one subject showed significant decrements on only two tasks (auditory vigilance and both measures of probability monitoring); two subjects showed decrements on three tasks and one subject showed decrements on four tasks. Although the overall pattern of performance for the HOPE-II crew and the second crew in OPN-360 was rather similar in the decrements shown over time, the average levels of performance of the HOPE-II crew were somewhat superior.

The HOPE-II crews experienced much less difficulty in remaining awake on duty; the most parsimonious explanation of this finding is that the group-performance tasks prevented gross deviations from alert performance. That is to say, the code-lock task revealed to the remainder of the crew (and the target identification task revealed to the commander) whether or not a given individual was at least alert enough to do his part on these group tasks. The pressures exerted by the group on the individual who tended to doze were abundantly evidenced. These pressures probably also tended to maintain at a higher level the general alertness of the individual, and this should explain in part the general superiority of the performance of the HOPE-II crew over that of the second OPN-360 crew.

The HOPE-III data present an entirely different picture. These subjects showed essentially no performance decrements. Specifically, three subjects showed significant decrements on auditory vigilance, and one of these three also showed decrements on the target-identification task under both conditions of presentation. Rather than decrements, the general picture for the HOPE-III subjects was one of improvements in performance over the course of 30 days. In addition to showing improvements, these subjects were superior to the HOPE-II subjects on all except the arithmetic task.

It seemed quite clear to the experimenters that the HOPE-II subjects were more highly motivated than those of HOPE-III, and that if the groups were not equal in ability, then it was the HOPE-II group that held a slight advantage over HOPE-III. Therefore, the differences in the performance of the two groups is apparently a direct reflection of the relative impacts of the two work-rest schedules. In other words, at least some highly motivated subjects are able to follow a 4-2 schedule without showing decrements; the extra 4 hours of performance per day, however, is achieved at a price. That price is the inability of the subjects to sustain performance at the level generally maintained by a comparable group working only 12 hours per day on a 4-4 schedule. Another possible component of the total price is being assessed in current studies; viz, the relative "performance reserves" of subjects on 4-2 versus 4-4 schedules when an additional stress (sleep deprivation) is imposed.

REFERENCES

1. Adams, O. S., Aircrew Fatigue Problems During Extended Endurance Flight, Phase I: Planning. WADC Technical Report 57-510, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, May 1958.

2. Adams, O. S. and Chiles, W. D., Human Performance as a Function of the Work-Rest Cycle. WADD Technical Report 60-248, Wright Air Development Division, Wright-Patterson Air Force Base, Ohio, March 1960.

3. Adams, O. S. and Chiles, W. D., Human Performance as a Function of the Work-Rest Ratio During Prolonged Confinement. ASD Technical Report 61-720, Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, November 1961.

4. Adams, O. S., Levine, R. B., and Chiles, W. D., Research to Investigate Factors Affecting Multiple-Task Psychomotor Performance. WADC Technical Report 59-120, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, March 1959.

5. Alluisi, E. A., Hall, T. J., and Chiles, W. D., Group Performance During Four-Hour Periods of Confinement. MRL Technical Documentary Report 62-70, 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, June 1962.

6. Ray, J. T., Martin, O. E., Jr., and Alluisi, E. A., Human Performance as a Function of the Work-Rest Cycle: A Review of Selected Studies. National Academy of Sciences — National Research Council Publication 882, 1961.

7. Rotta, A., Cánepa, A., Hurtoda, A., Velásquez, T., and Chávez, R., Pulmonary Circulation at Sea Level and at High Altitudes, Journal of Applied Physiology, Vol. 9, pp. 328 - 336, 1956.

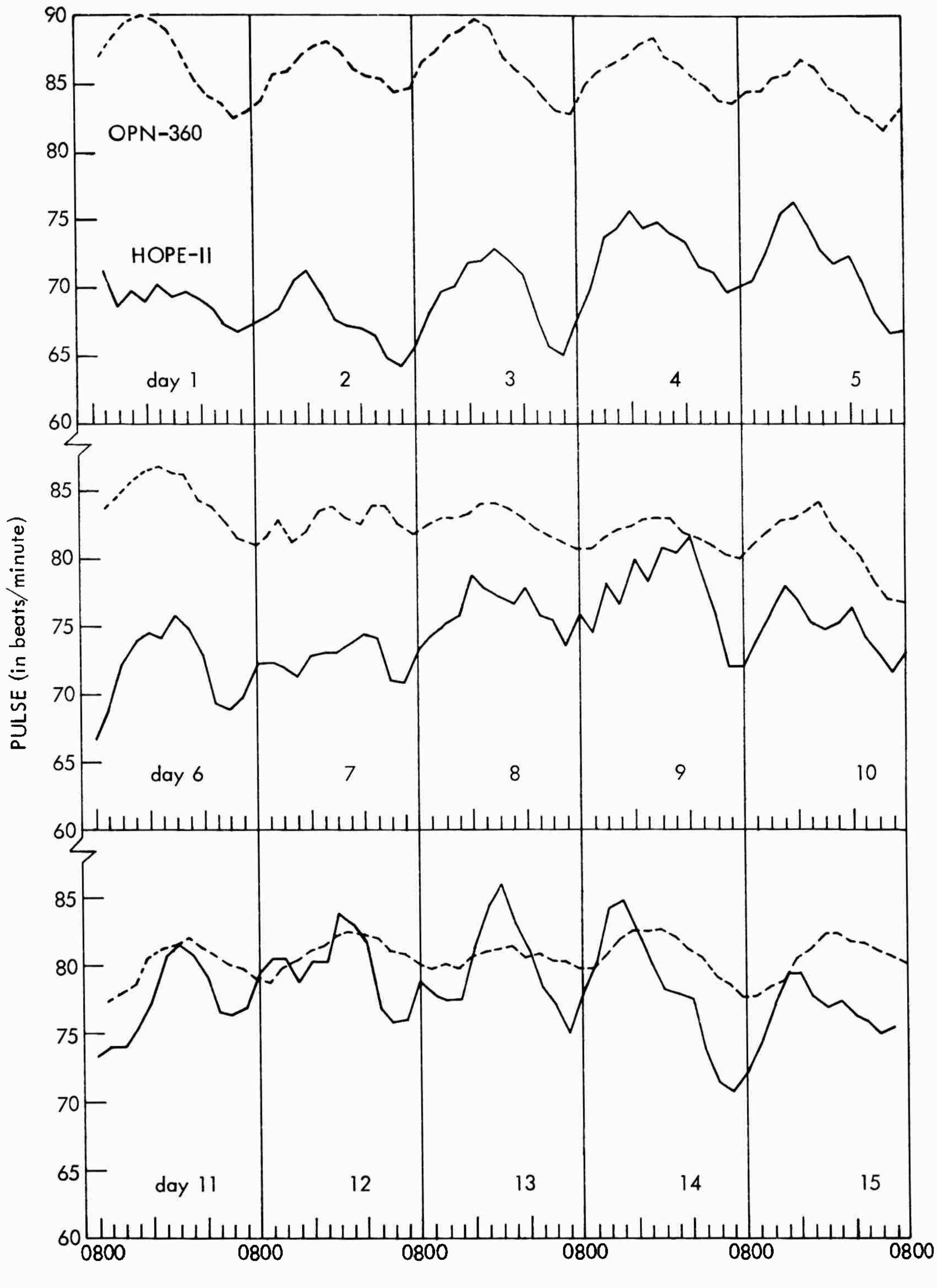

Figure 2. Mean pulse-rate data of OPN-360 and HOPE-II.

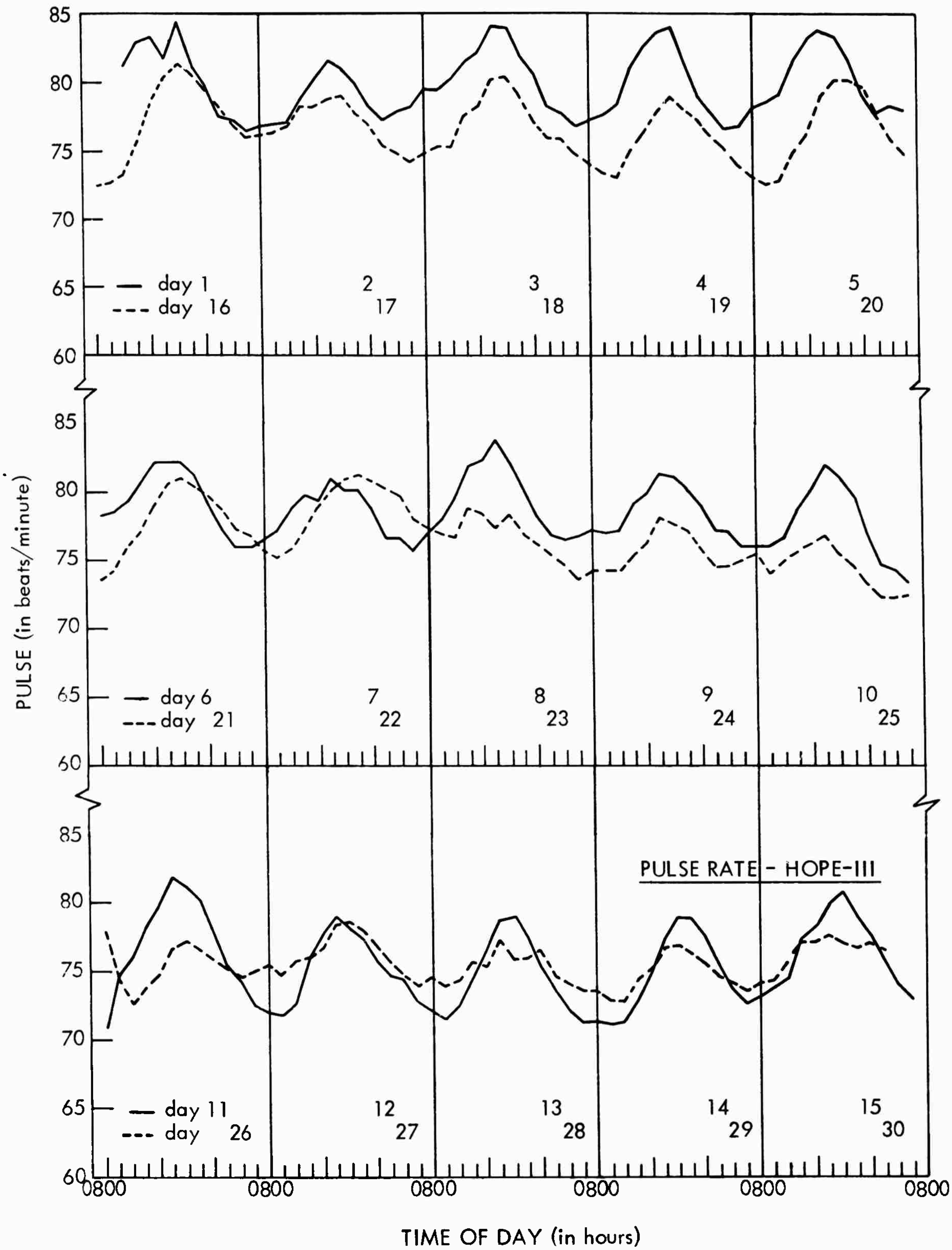

Figure 3. Mean pulse-rate data of HOPE-III.

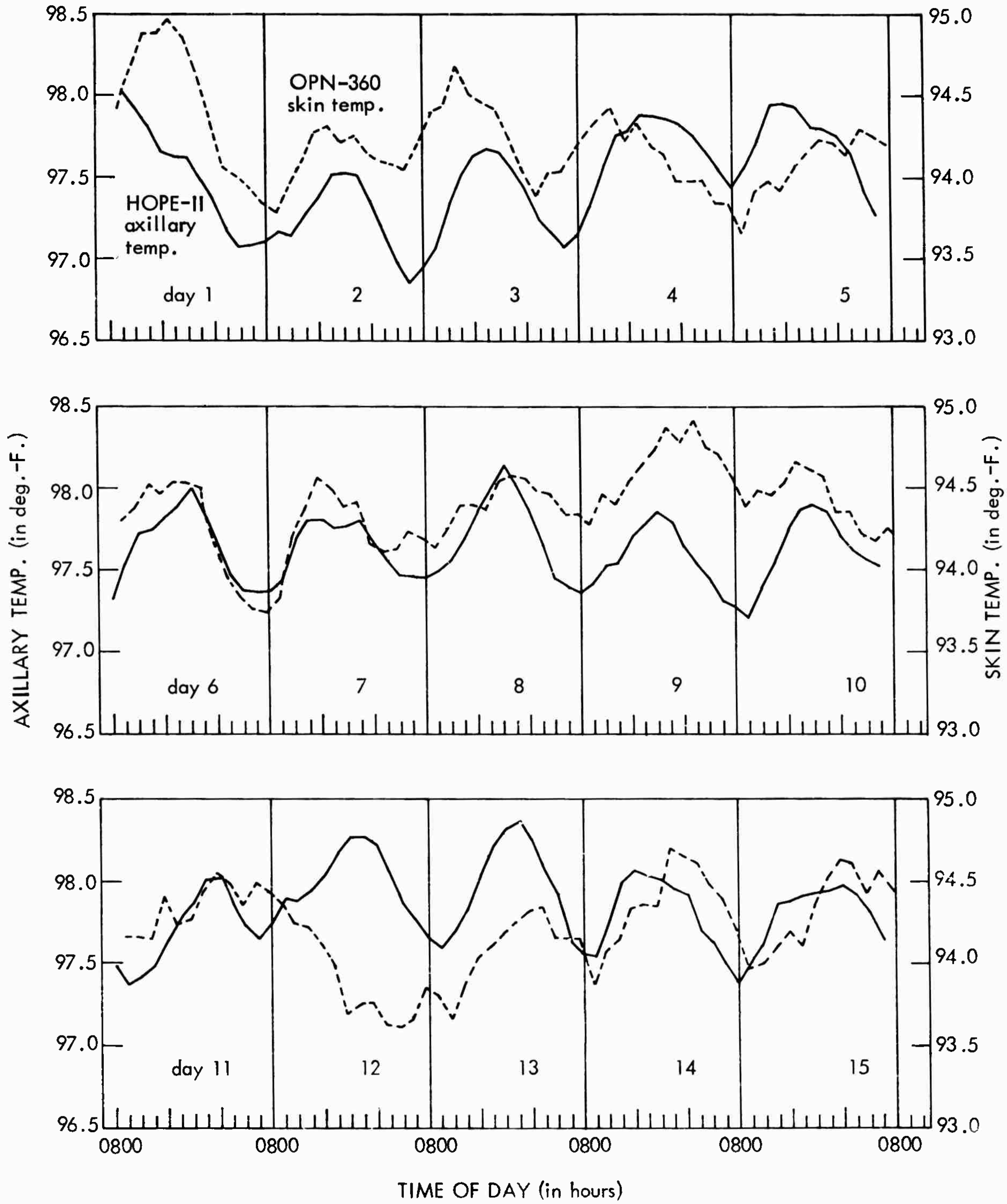

Figure 4. Mean temperature data of OPN-360 and HOPE-II.

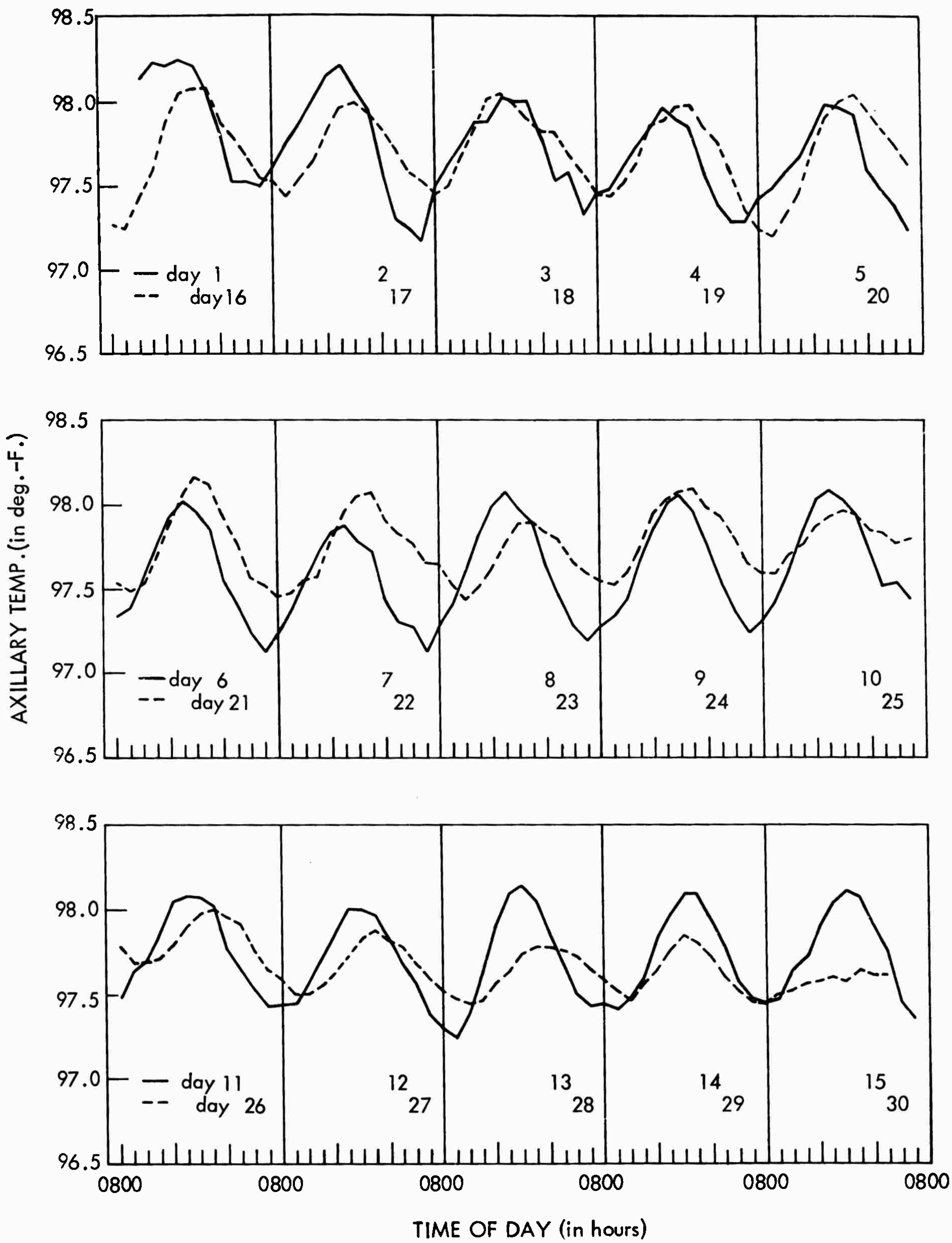

Figure 5. Mean temperature data of HOPE-III.

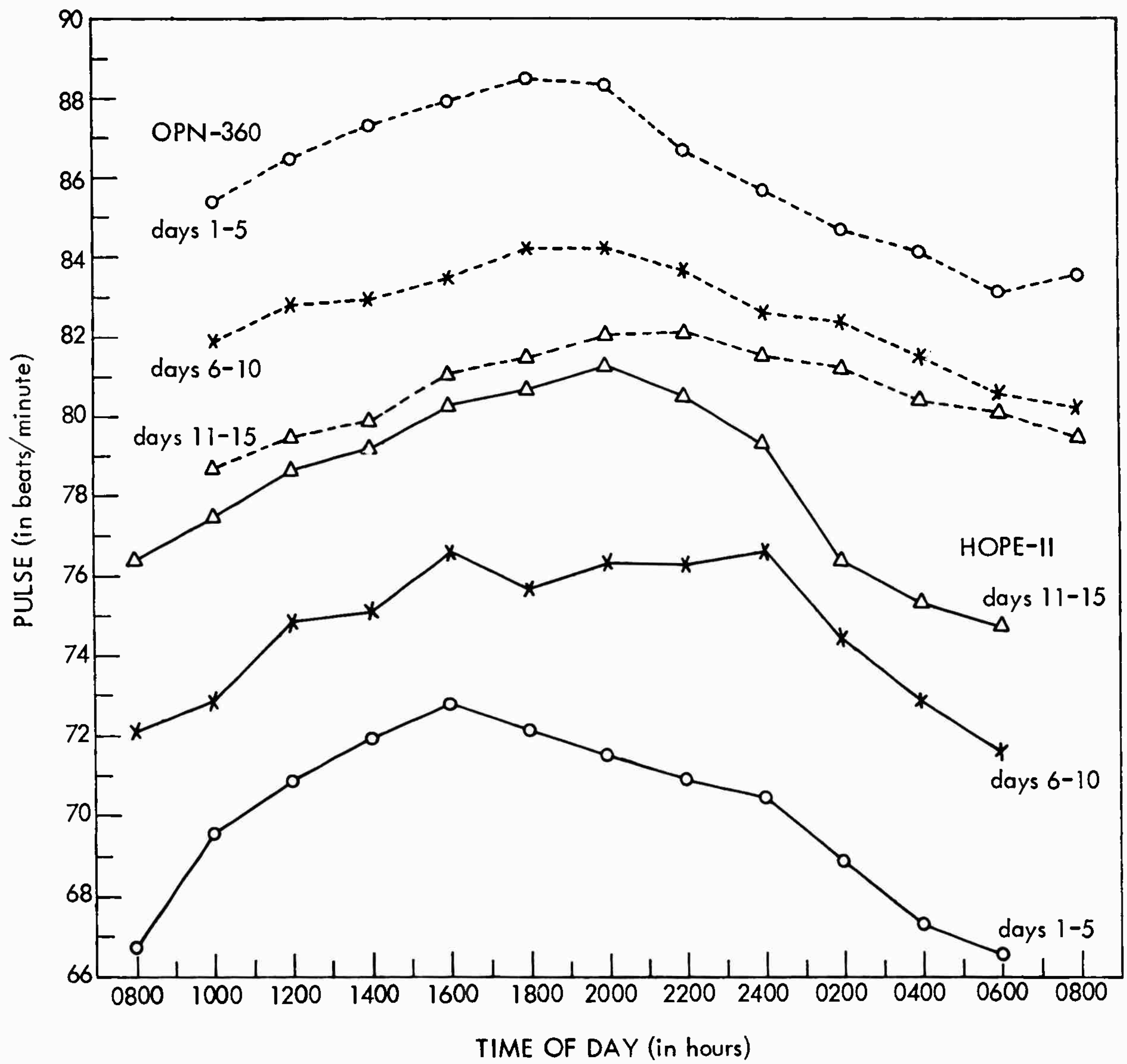

Figure 6. Within-day changes in pulse rate of OPN-360 and HOPE-II.

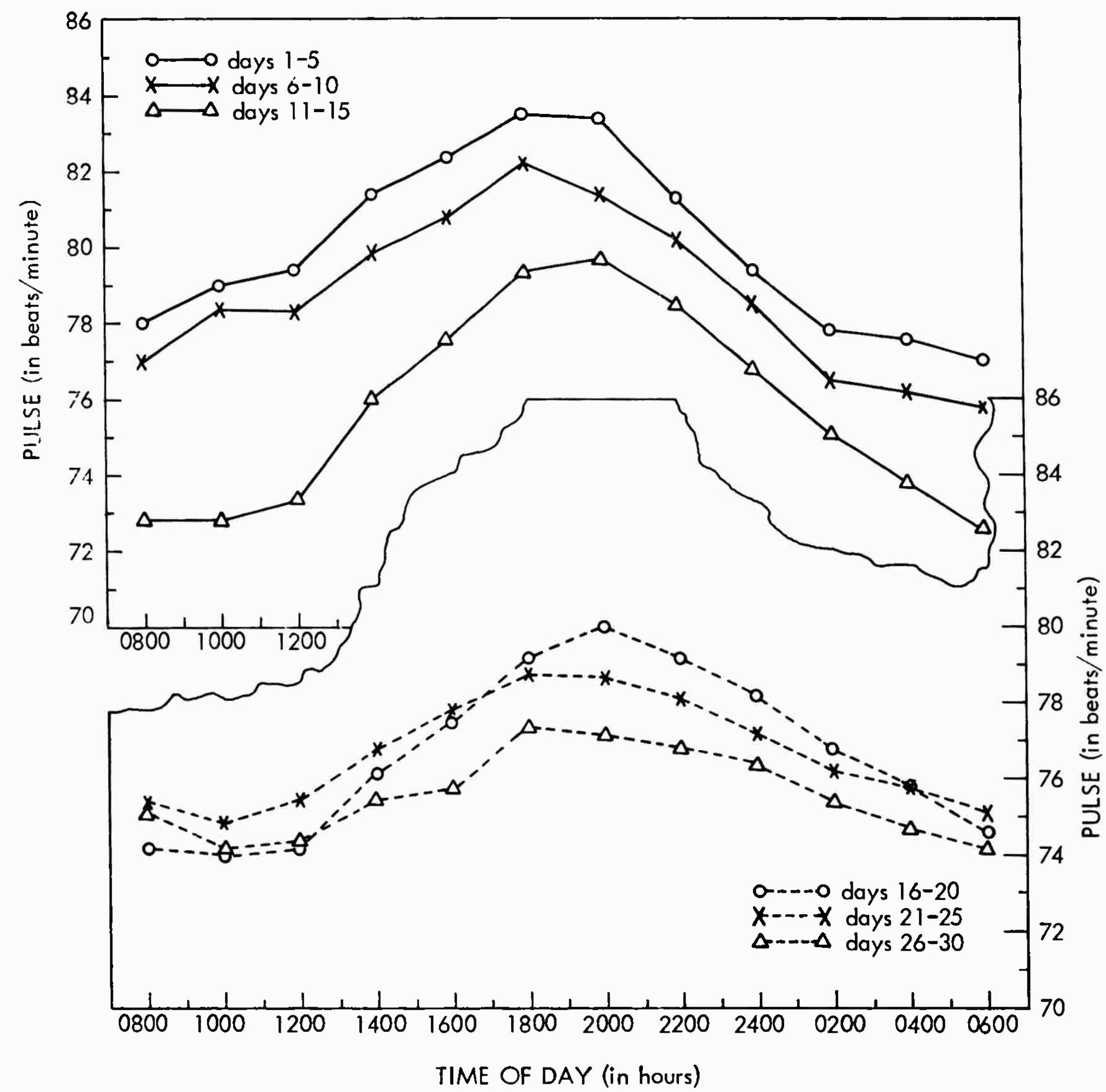

Figure 7. Within-day changes in pulse rate of HOPE-III.

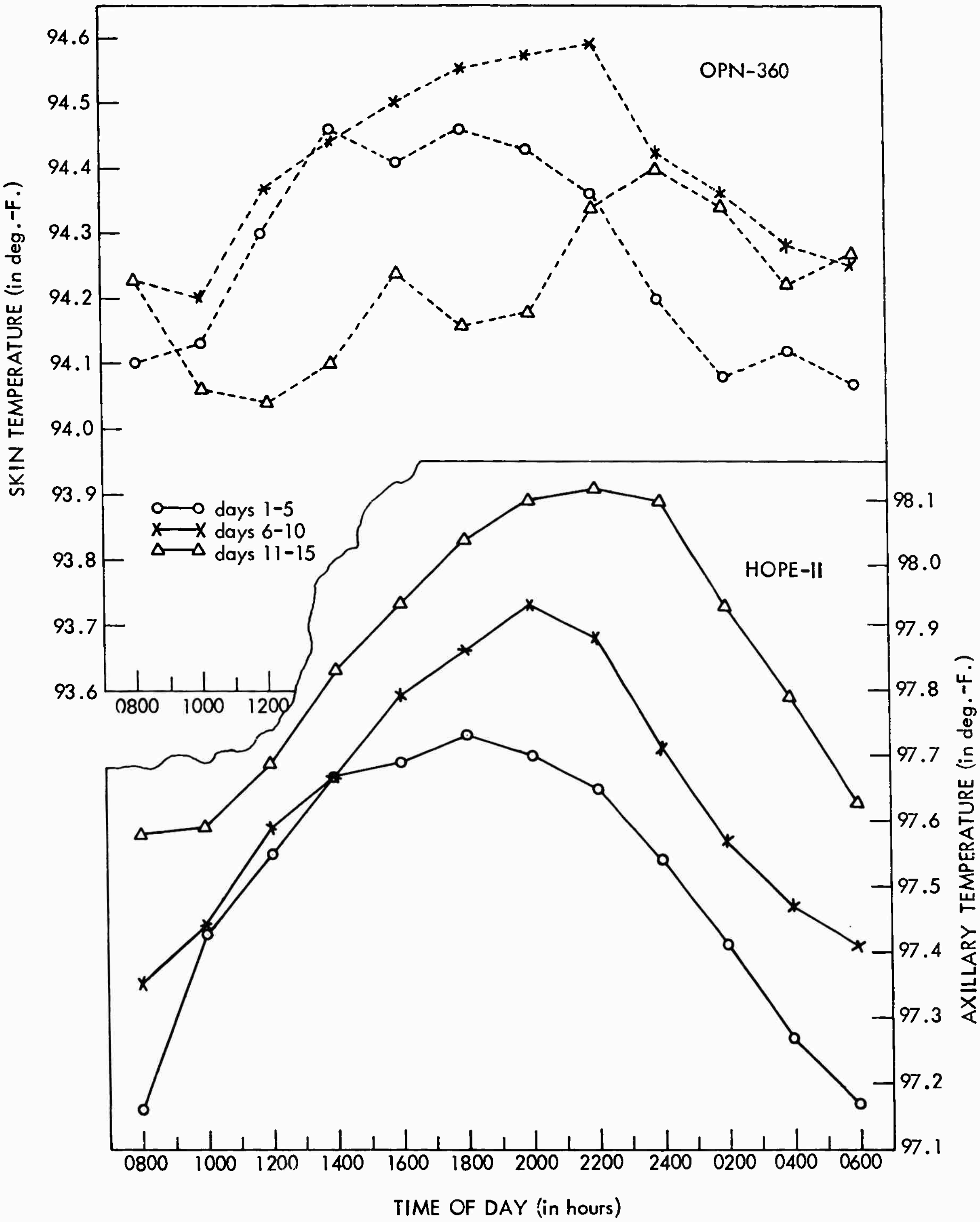

Figure 8. Within-day changes in temperature of OPN-360 and HOPE-II.

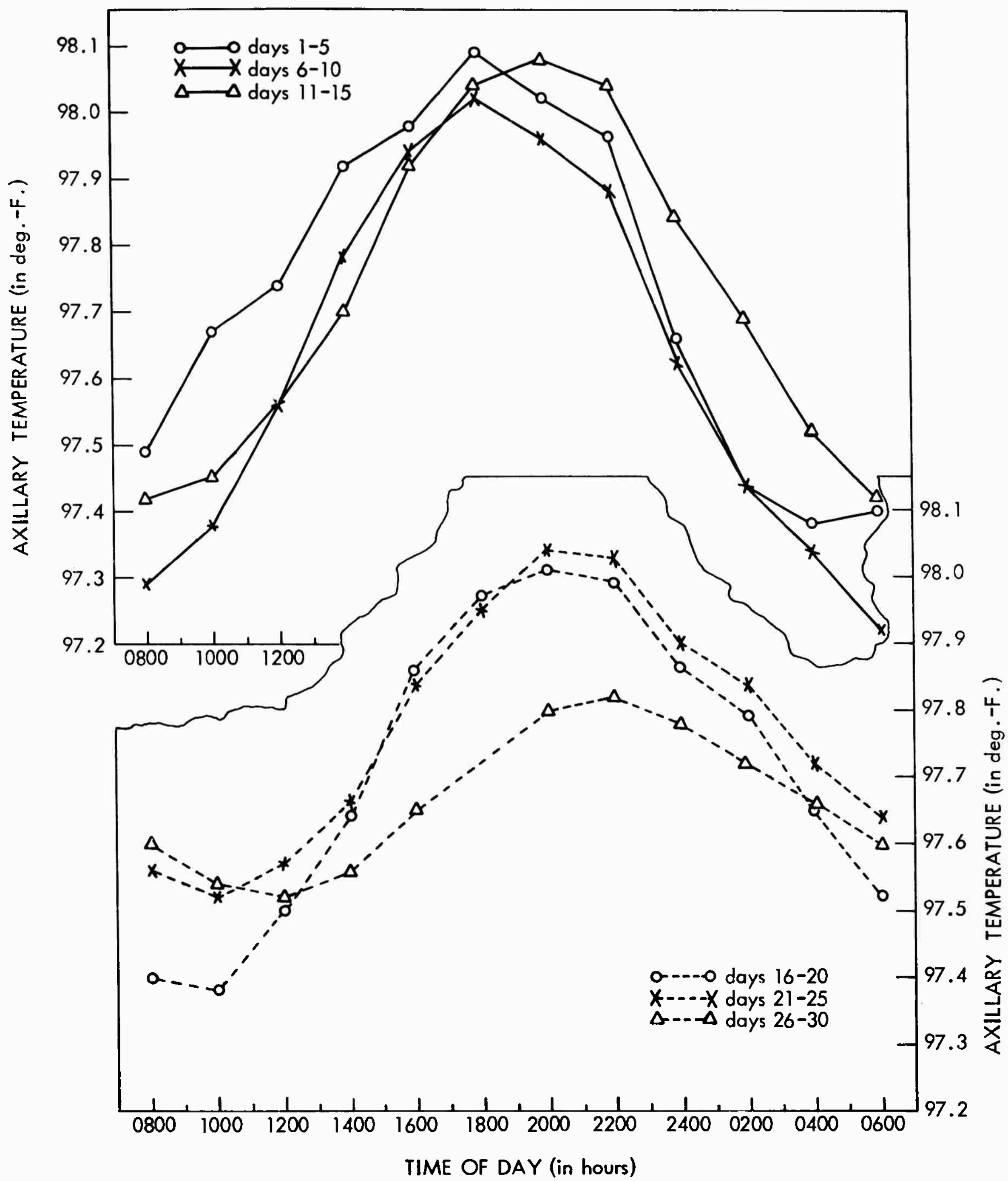

Figure 9. Within-day changes in axillary temperature of HOPE-III.

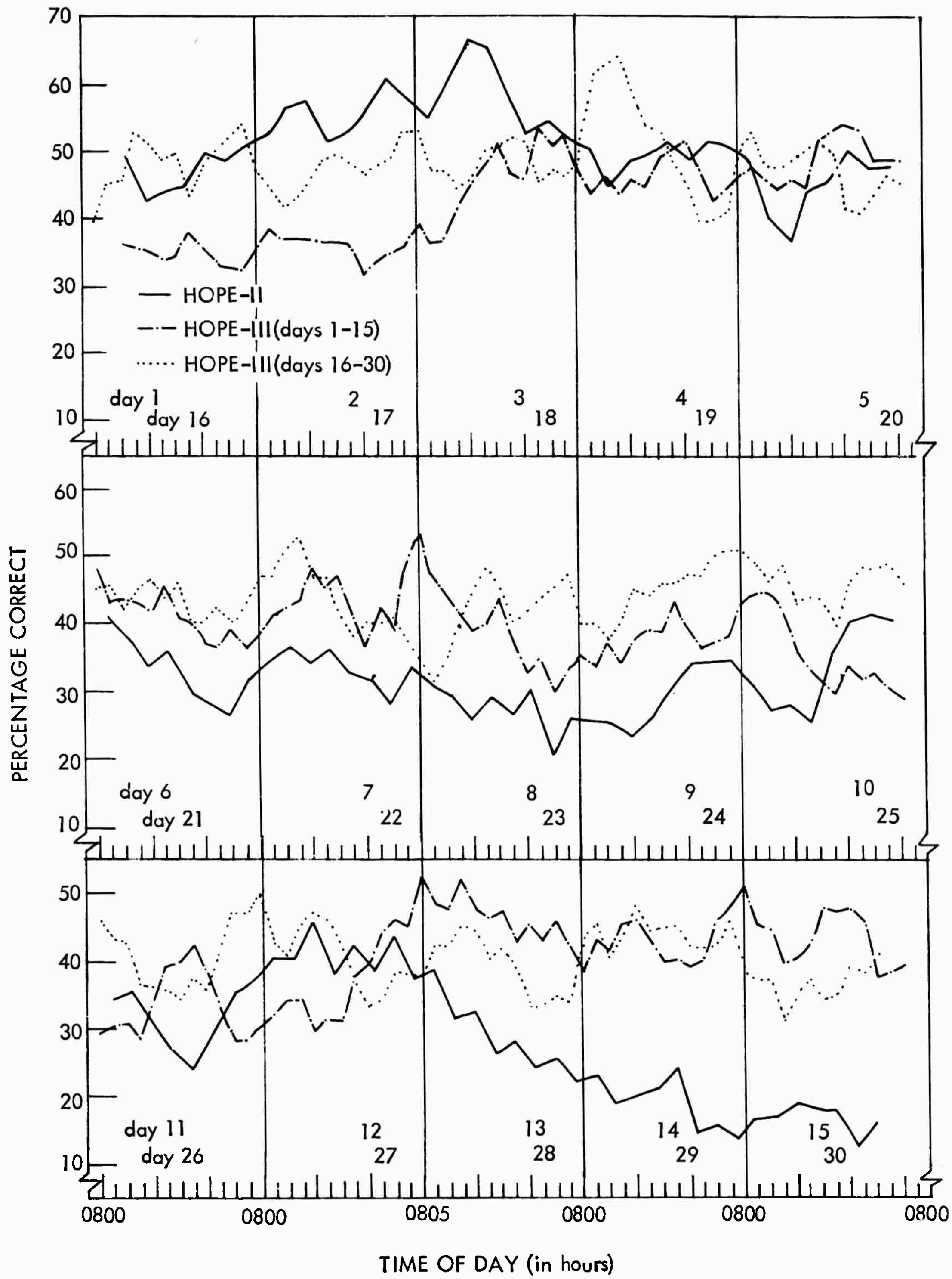

Figure 10. Mean percentages of correct auditory-vigilance signal detections: HOPE-II and HOPE-III.

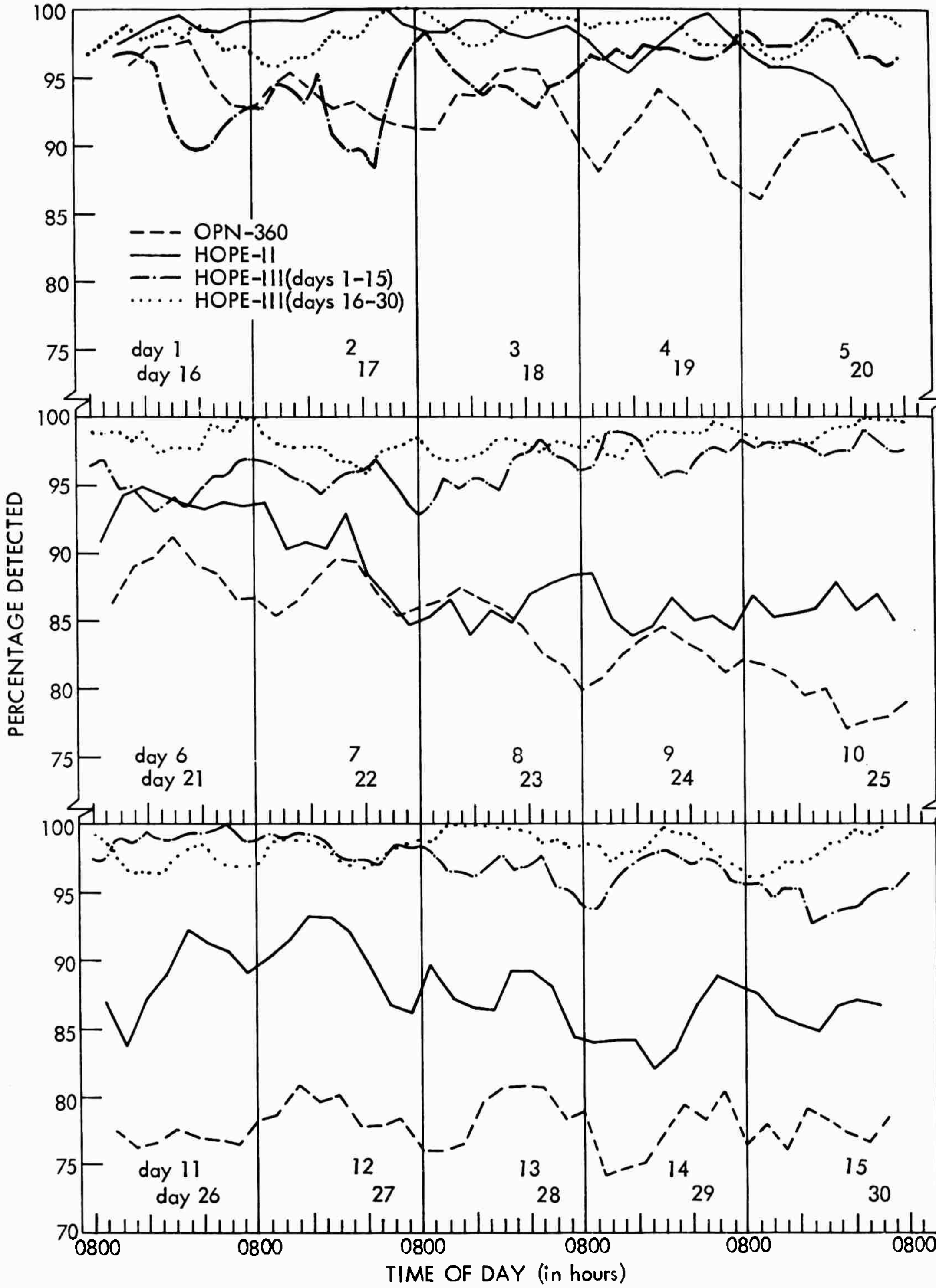

Figure 11. Mean percentages of correct probability-monitoring signal detections: OPN-360, HOPE II, and HOPE III.

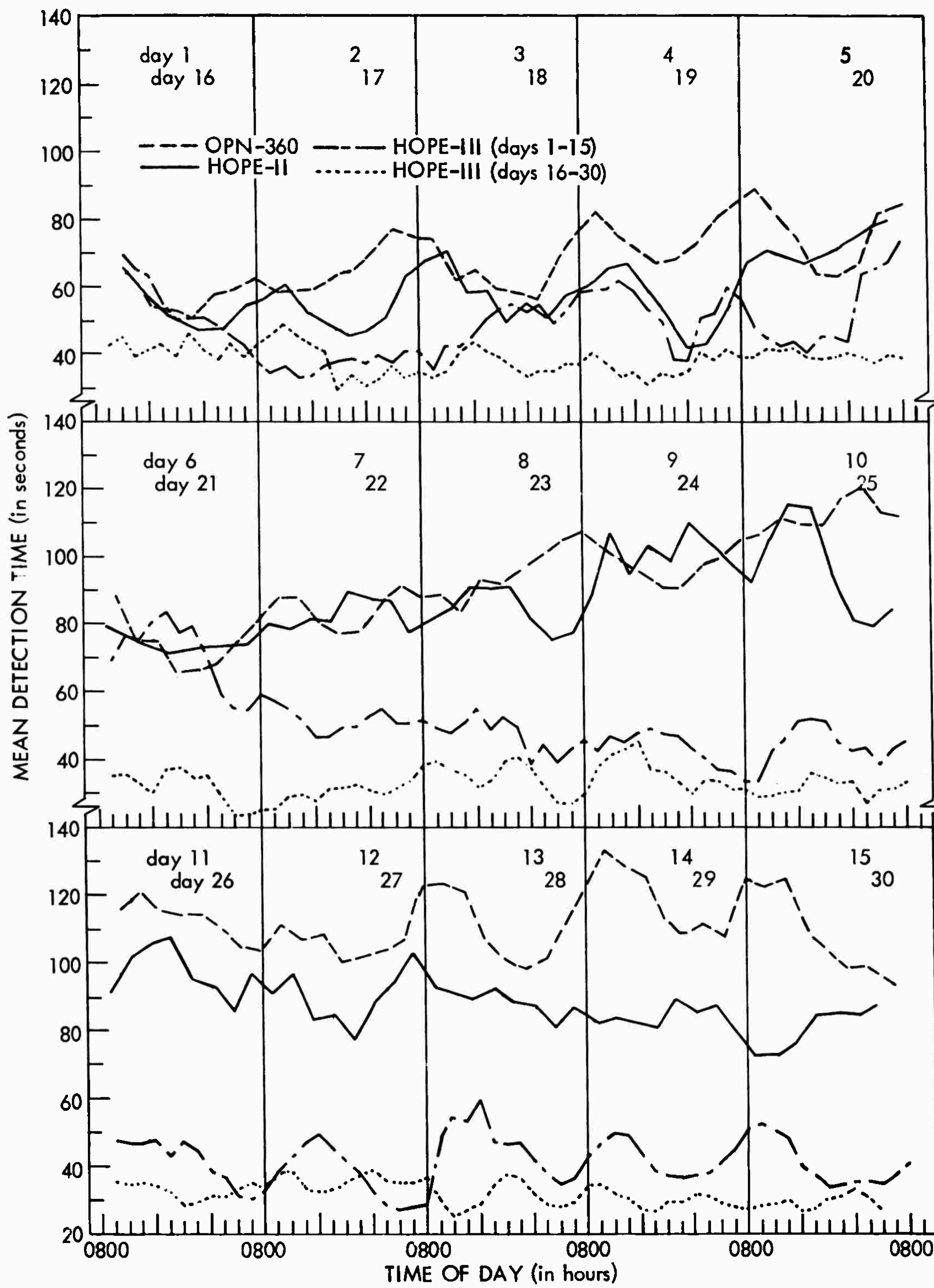

Figure 12. Mean time (in seconds) to detect probability-monitoring signal: OPN-360, HOPE-II, and HOPE-III.

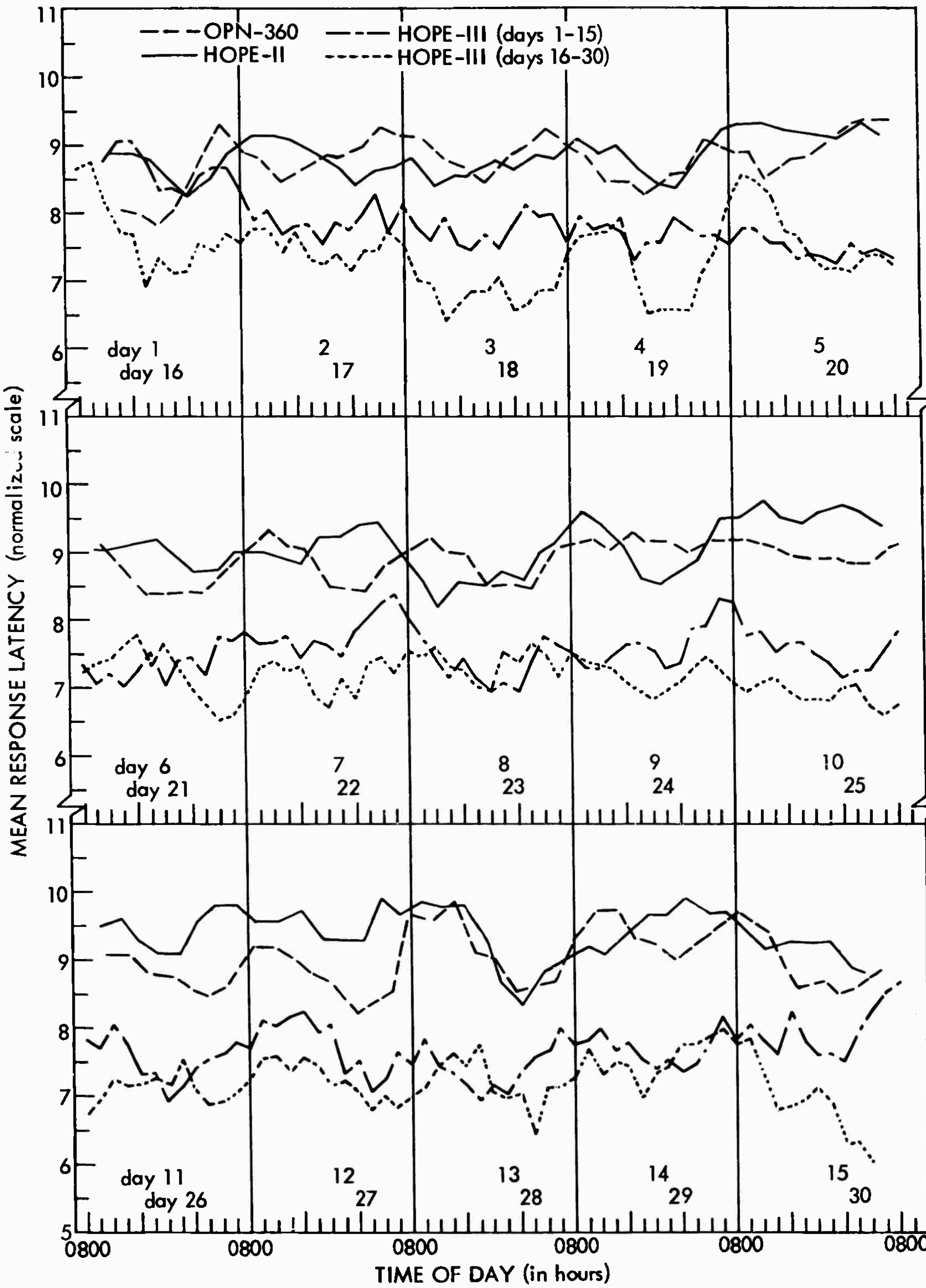

Figure 13. Mean response latency (normalized scale) in detecting red warning-light signals: OPN-360, HOPE-II, and HOPE-III.

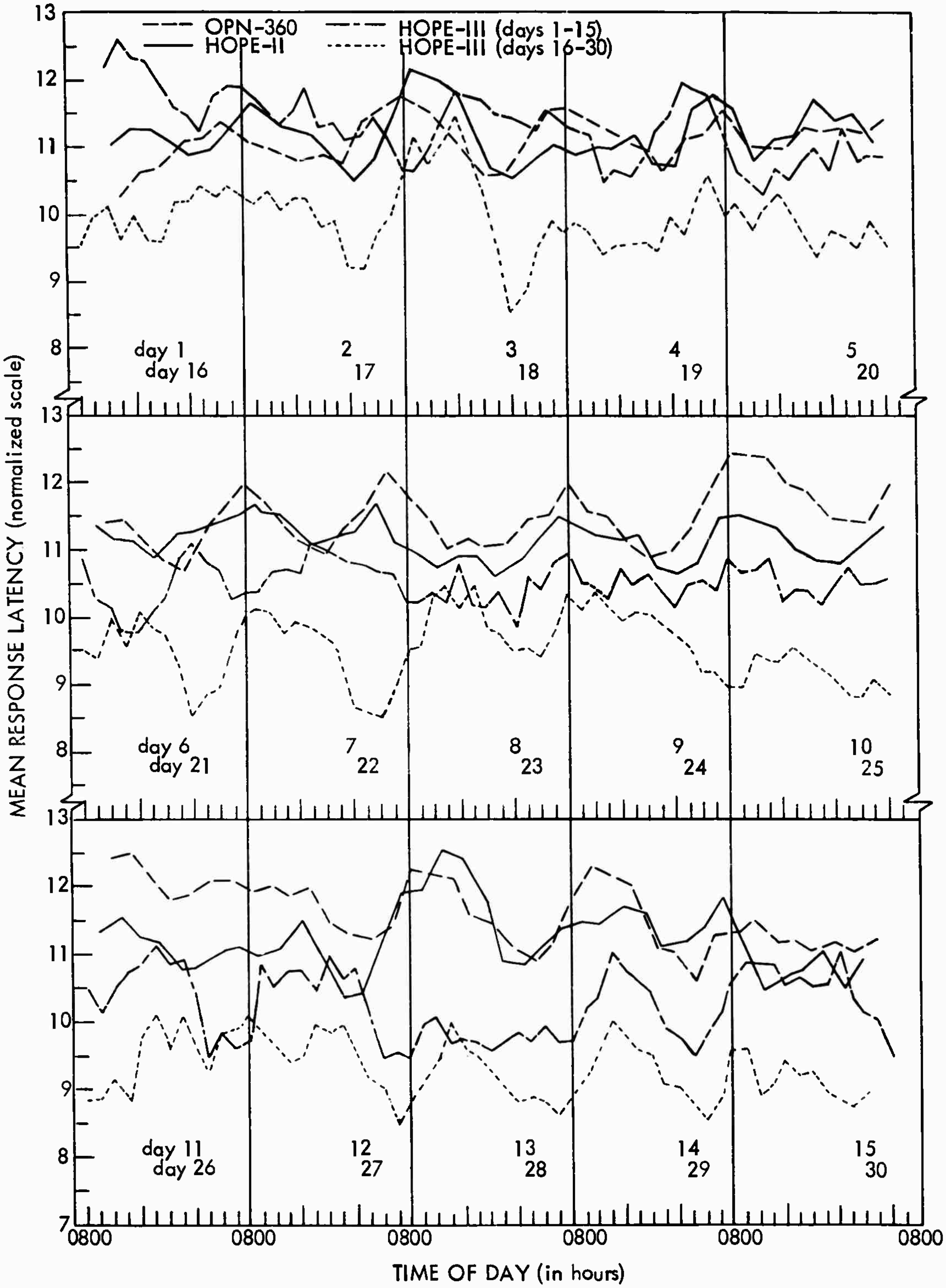

Figure 14. Mean response latency (normalized scale) in detecting green warning-light signals: OPN-360, HOPE-II, and HOPE-III.

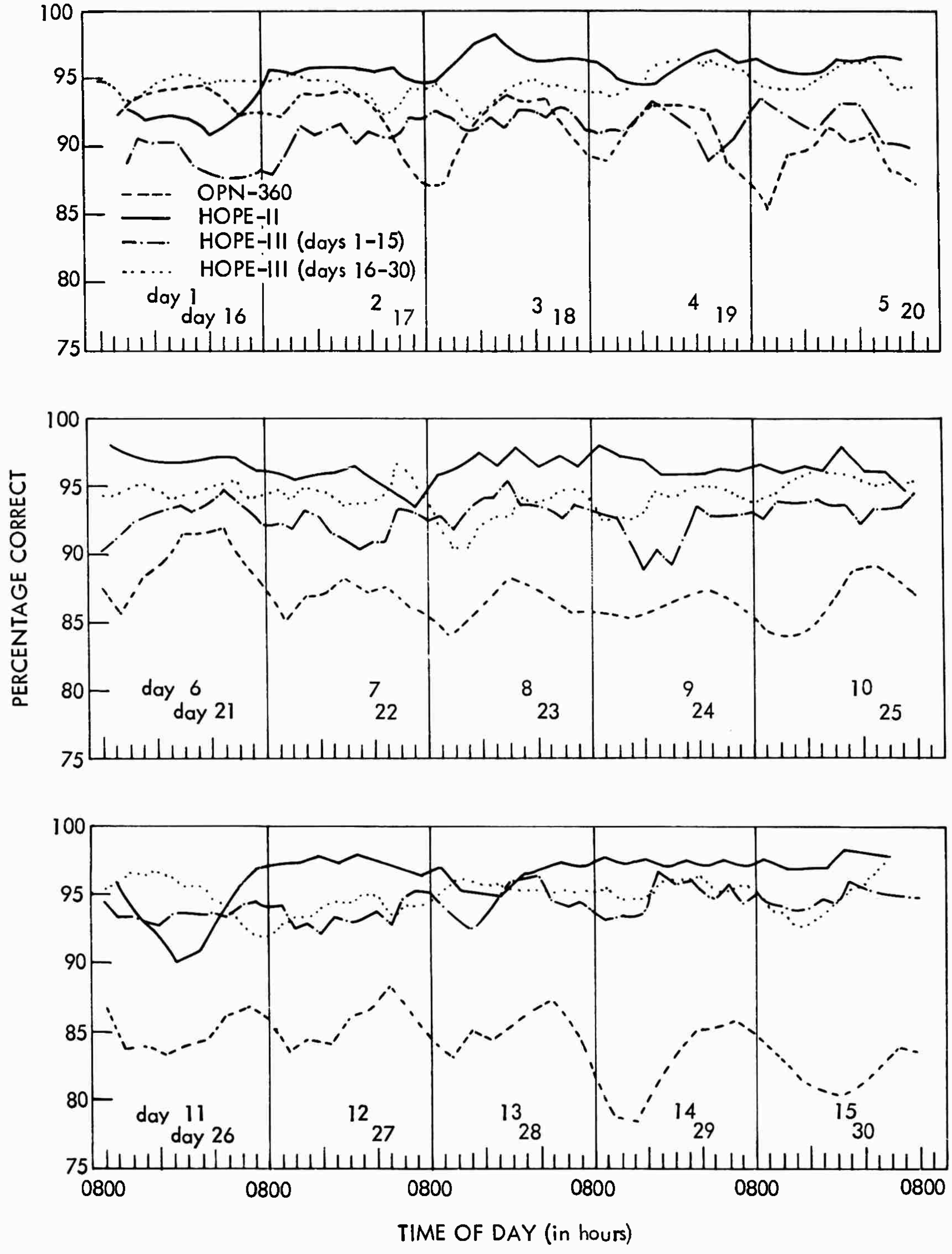

Figure 15. Mean percentages of correct arithmetic computations: OPN-360, HOPE-II, and HOPE-III, without simultaneous code-lock problems.

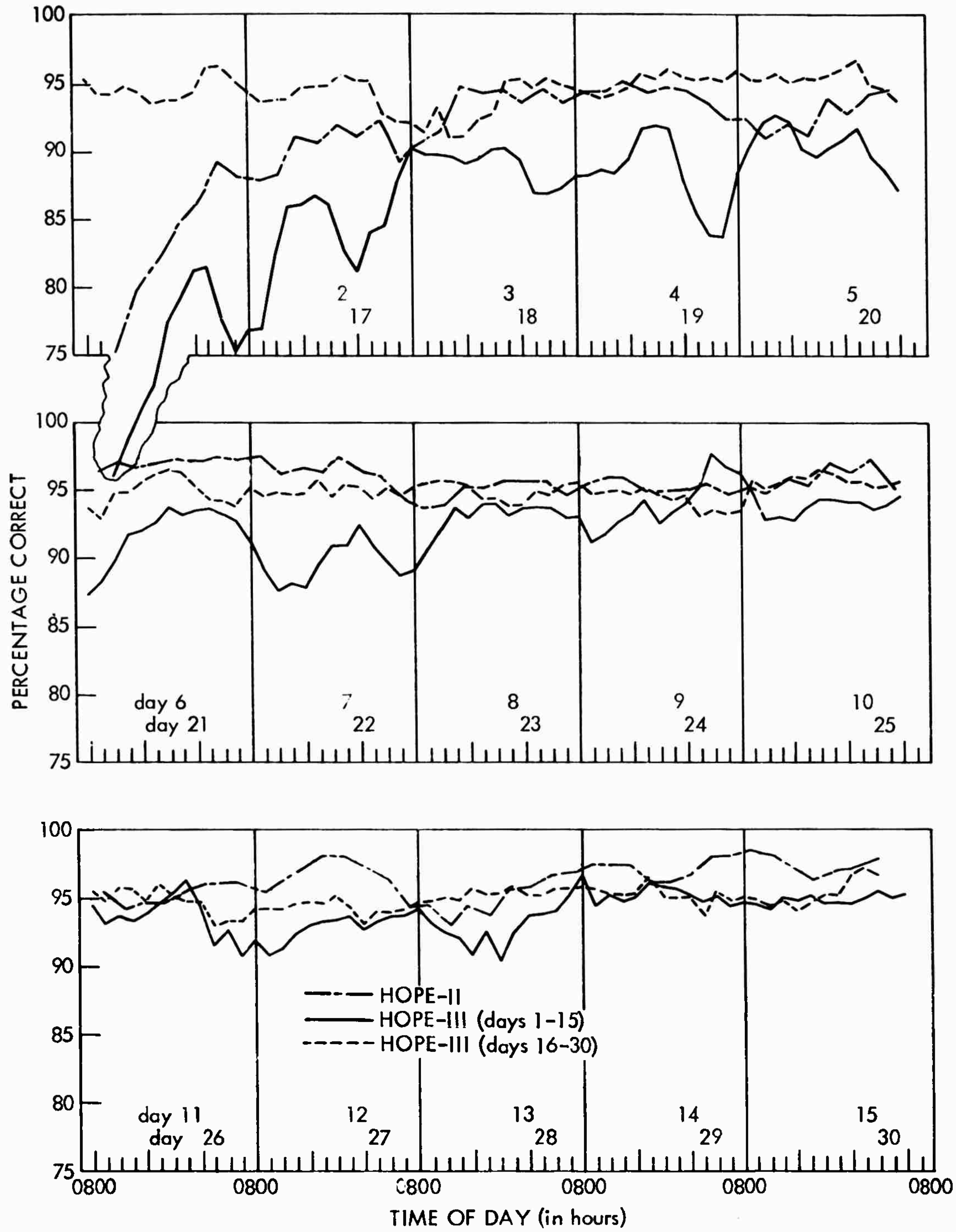

Figure 16. Mean percentages of correct arithmetic computations: HOPE-II and HOPE-III, with simultaneous code-lock problems.

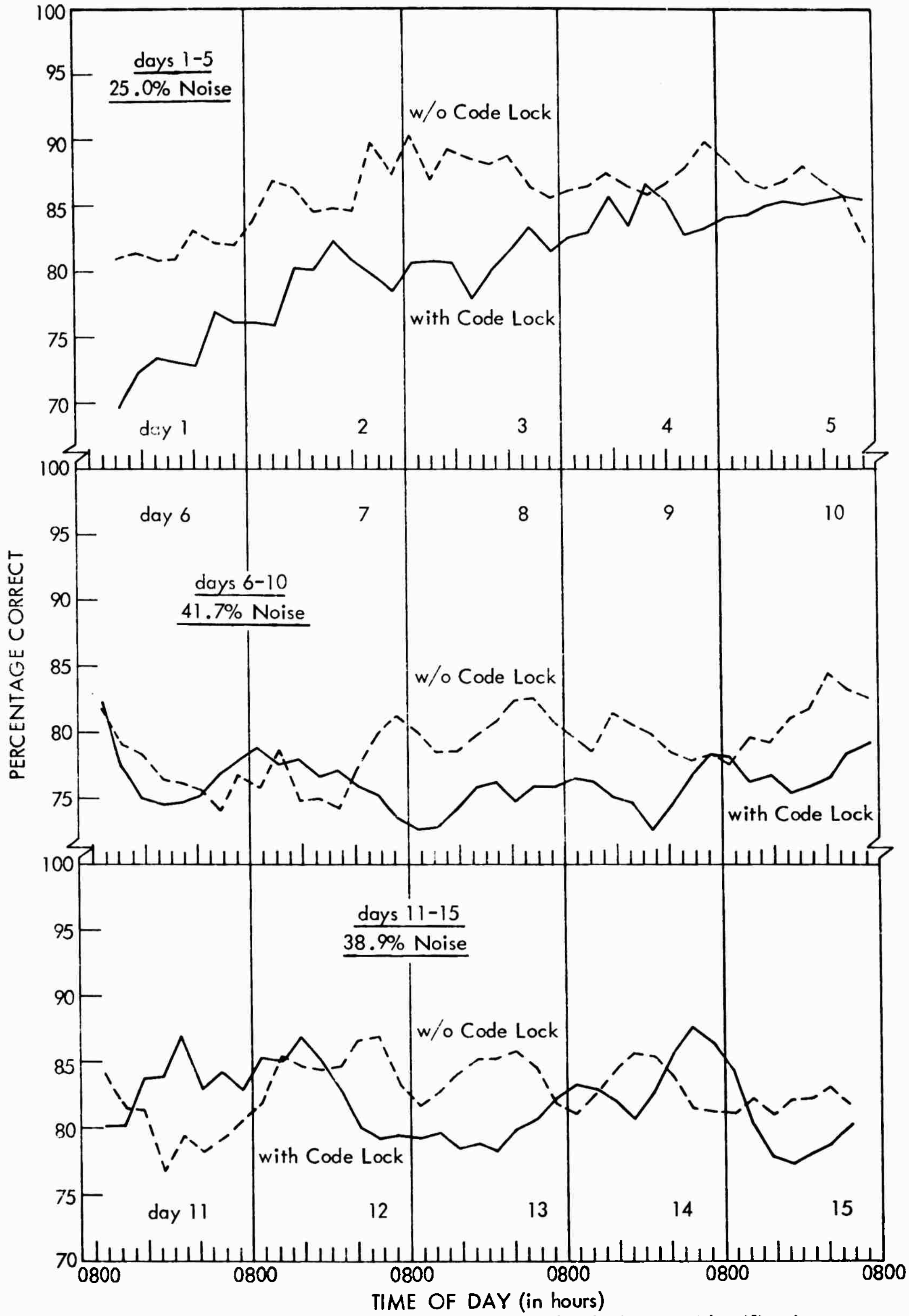

Figure 17. Mean percentages of correct individual target identification: HOPE-II, with and without simultaneous code-lock problems.

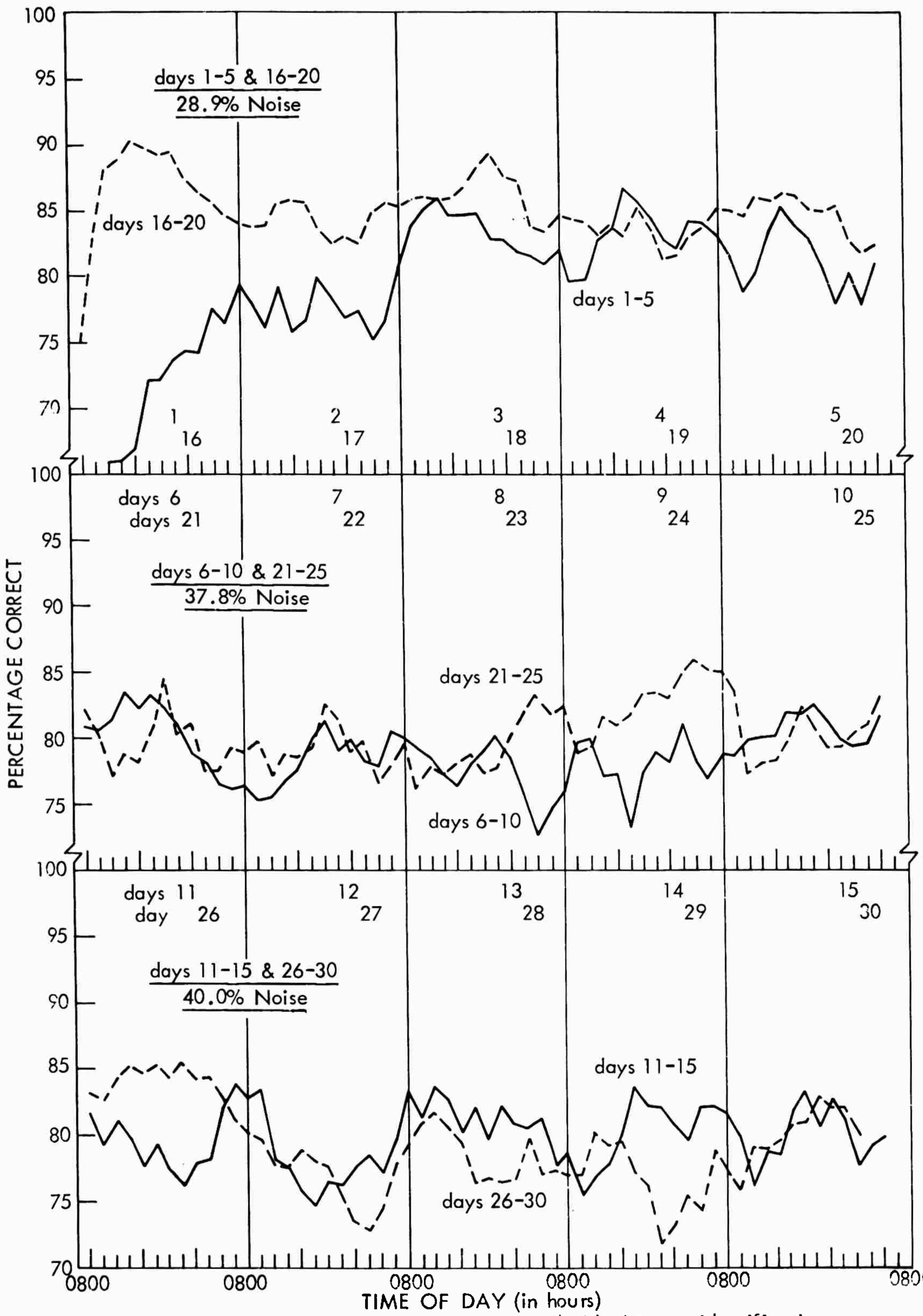

Figure 18. Mean percentages of correct individual target identification: HOPE-III, without simultaneous code-lock problems.

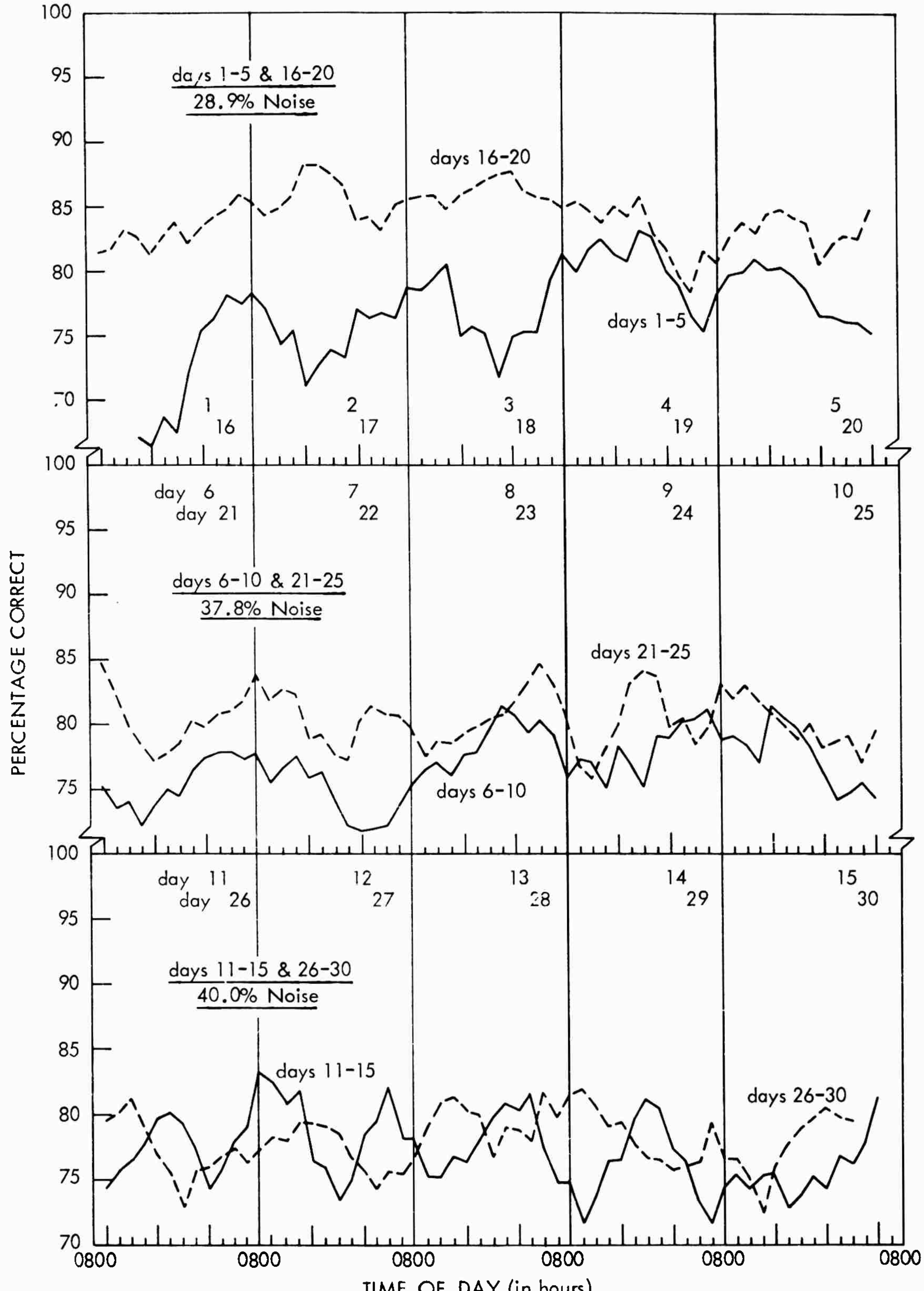

Figure 19. Mean percentages of correct individual target identification: HOPE-III, with simultaneous code-lock problems.

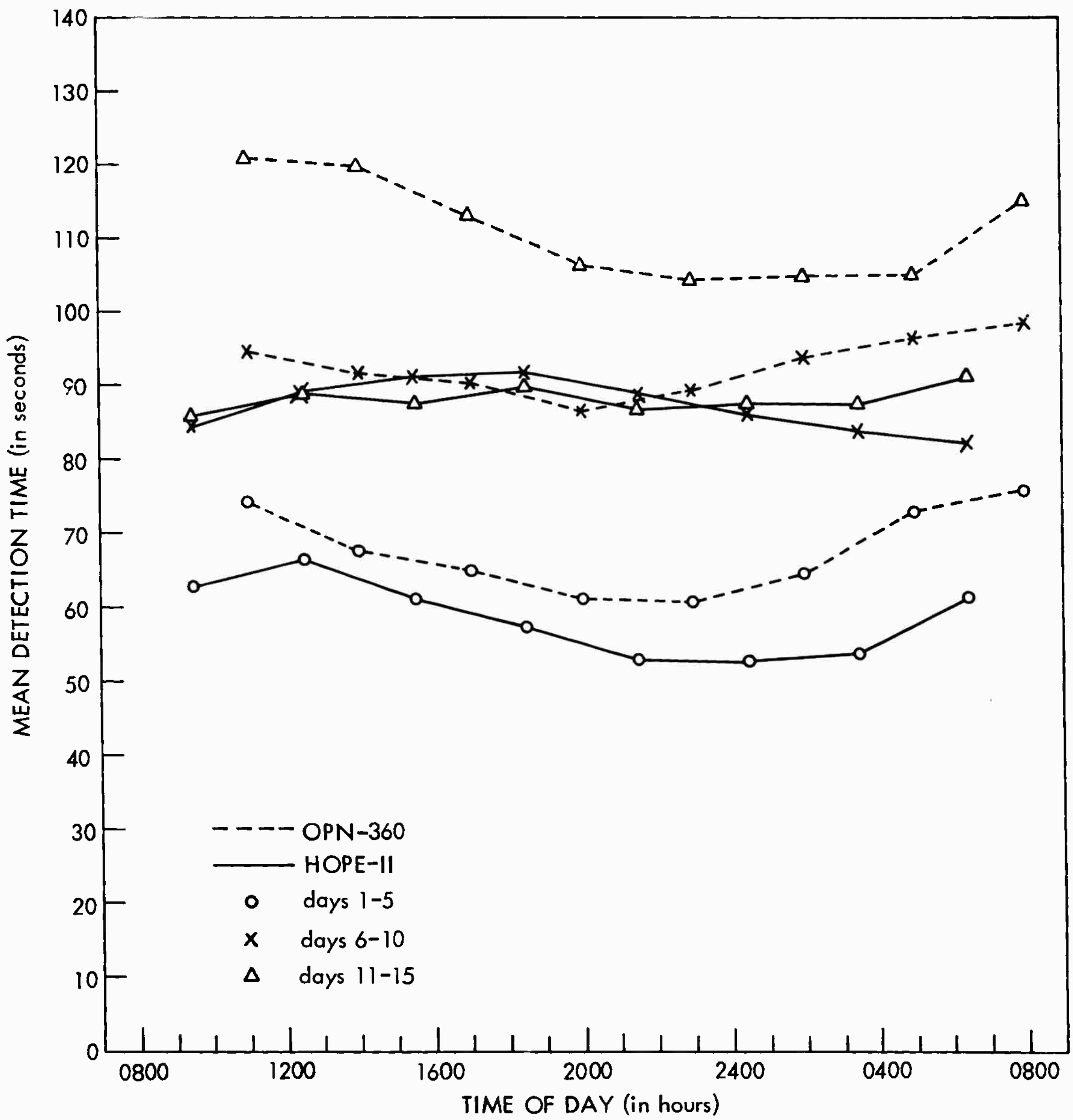

Figure 20. Within-day changes in probability-monitoring detection-time: OPN-360 and HOPE-II.

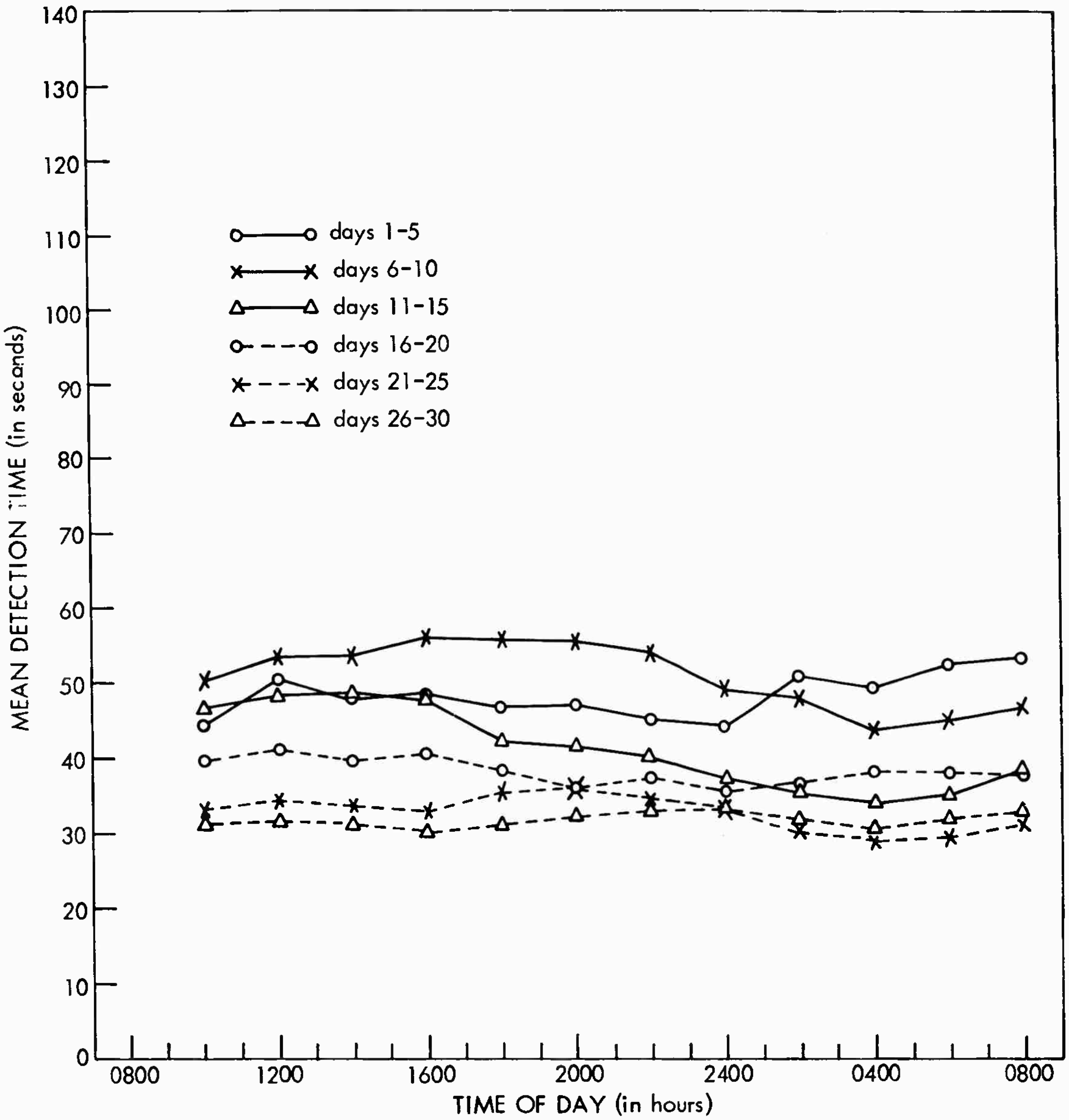

Figure 21. Within-day changes in probability-monitoring detection-time: HOPE-III.

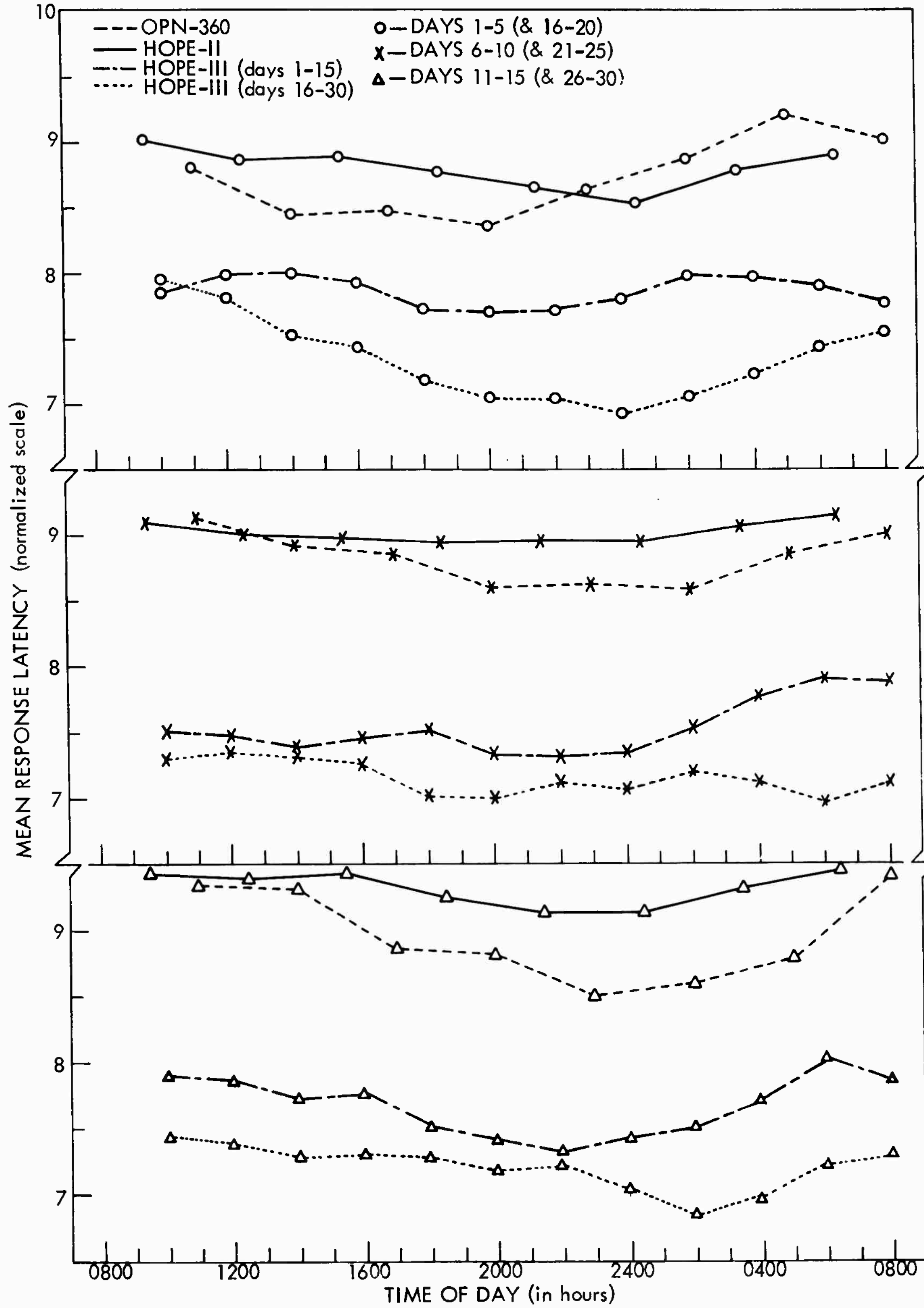

Figure 22. Within-day changes in response latency to red warning-lights

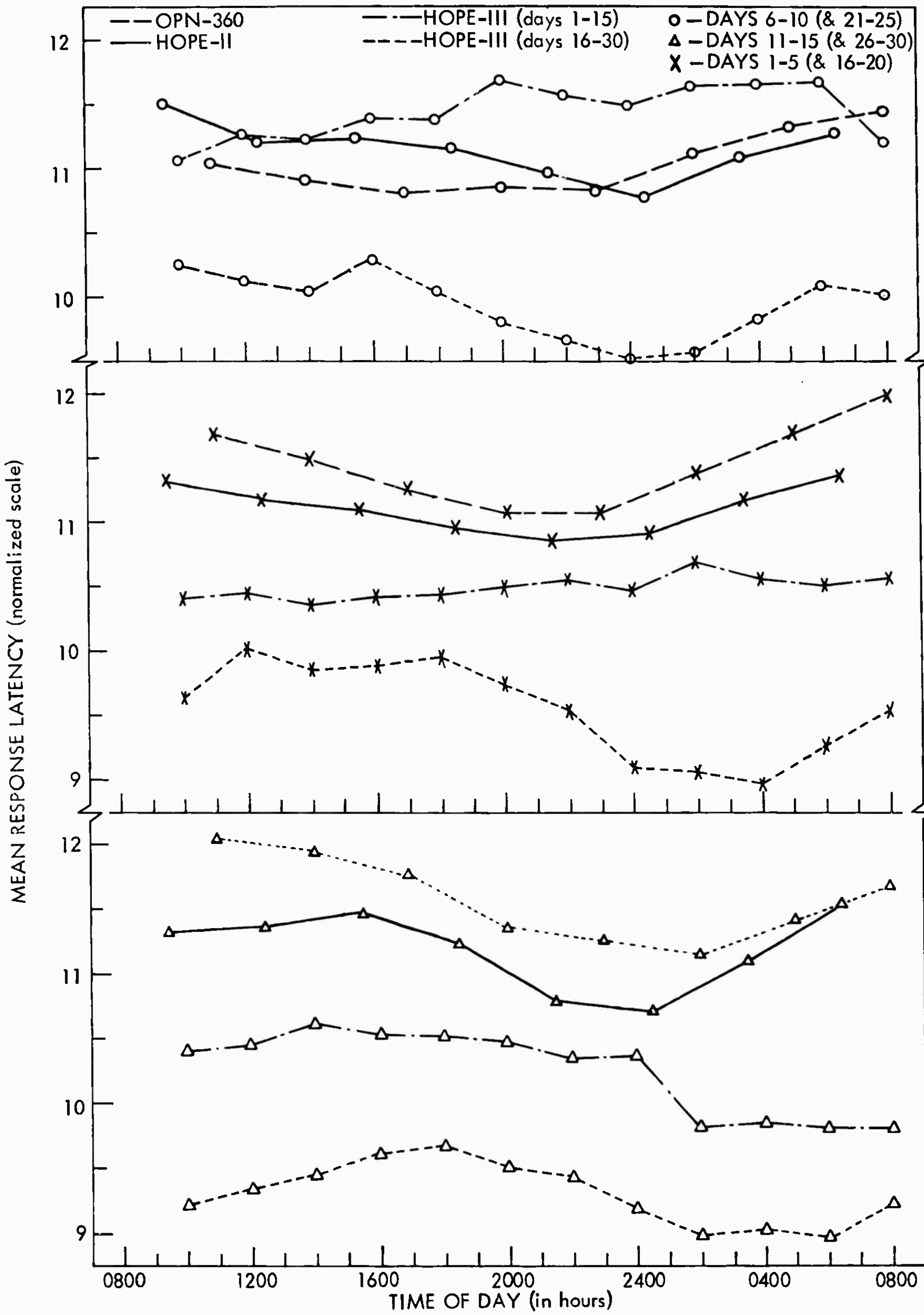

Figure 23. Within-day changes in response latency to green warning-lights

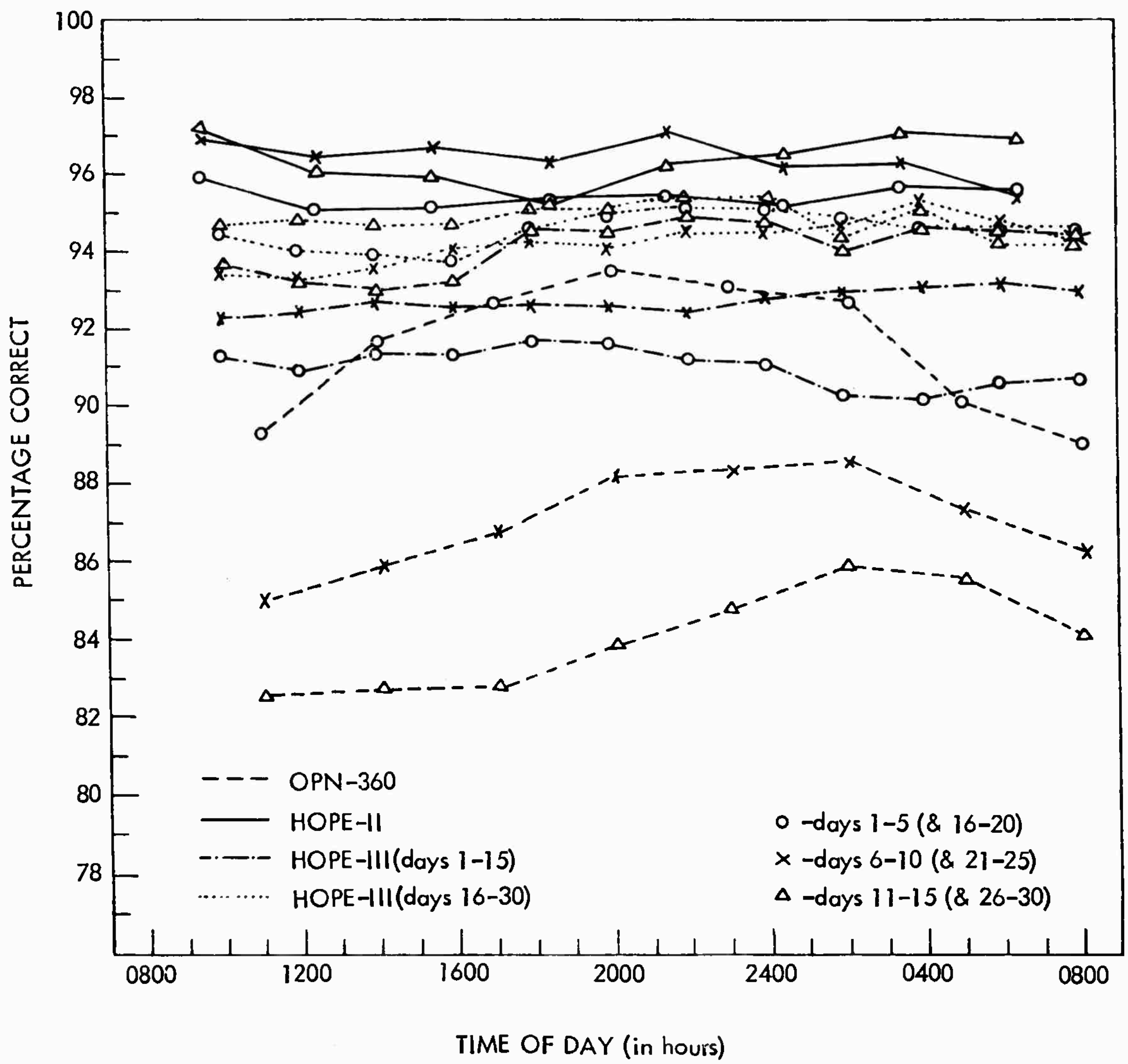

Figure 24. Within-day changes in arithmetic computations without simultaneous presentations of code-lock problems: OPN-360, HOPE-II, and HOPE-III.

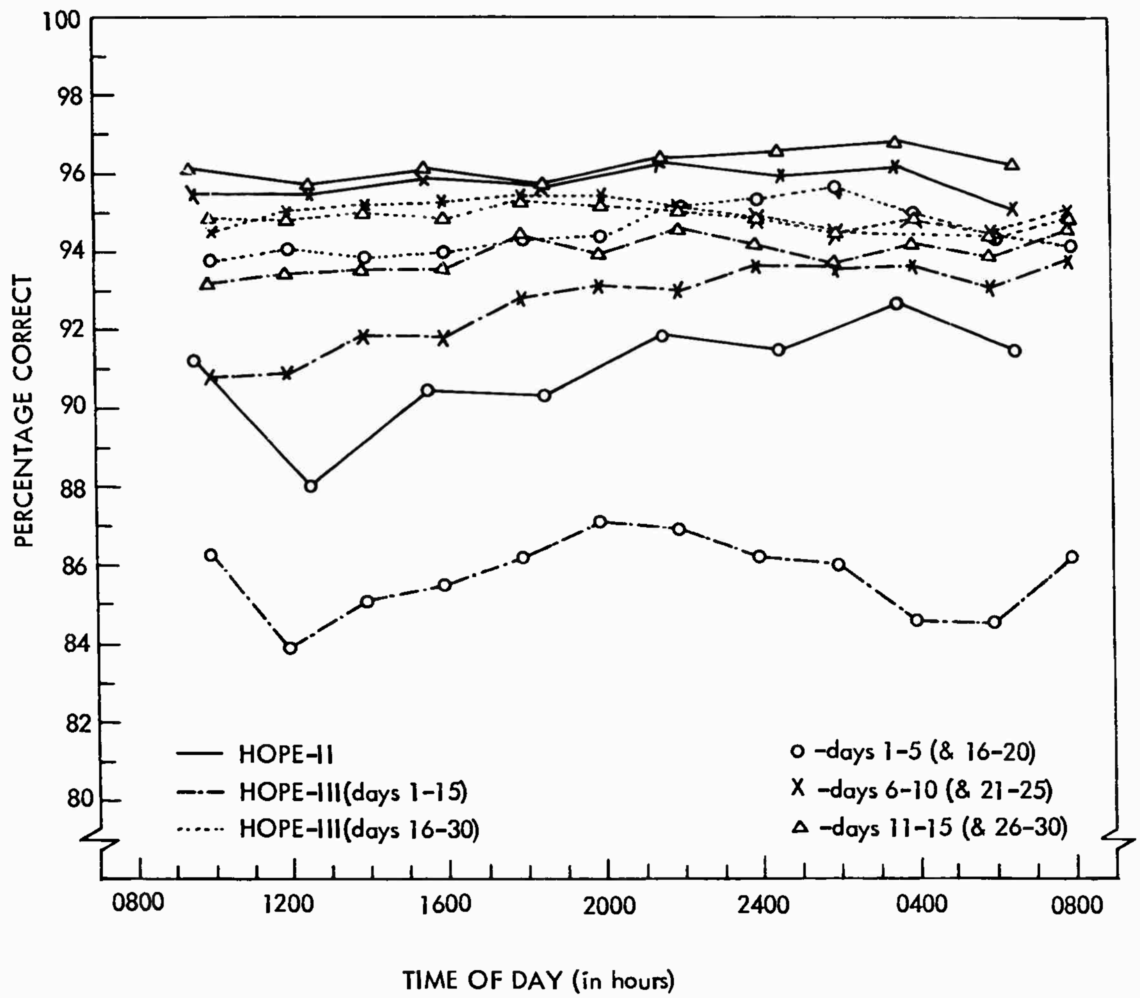

Figure 25. Within-day changes in arithmetic computations with simultaneous presentations of code-lock problems: HOPE-II and HOPE-III.

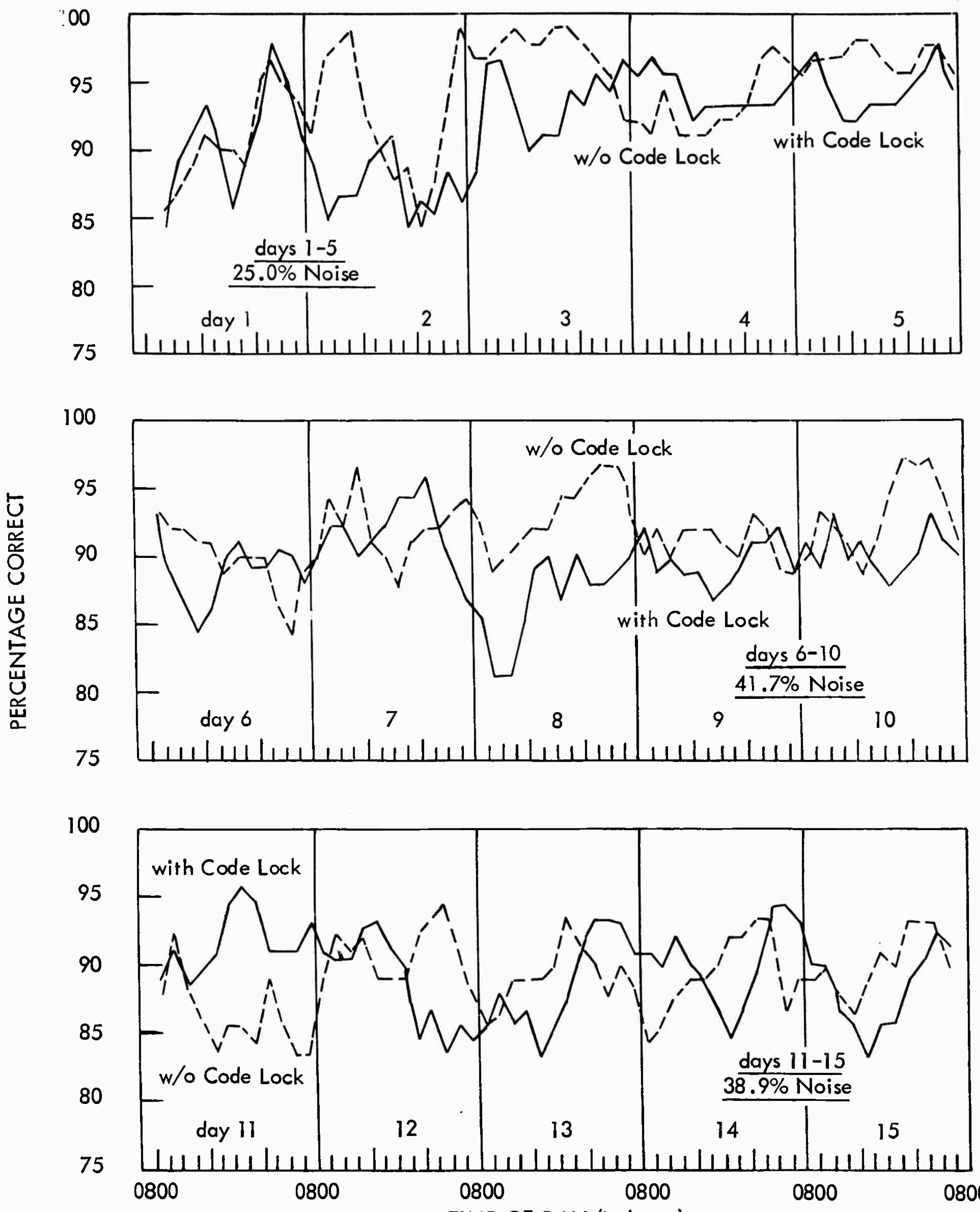

Figure 26. Mean percentages of correct commander's final decisions in target identification with and without simultaneous presentation of code-lock problems: HOPE-II.

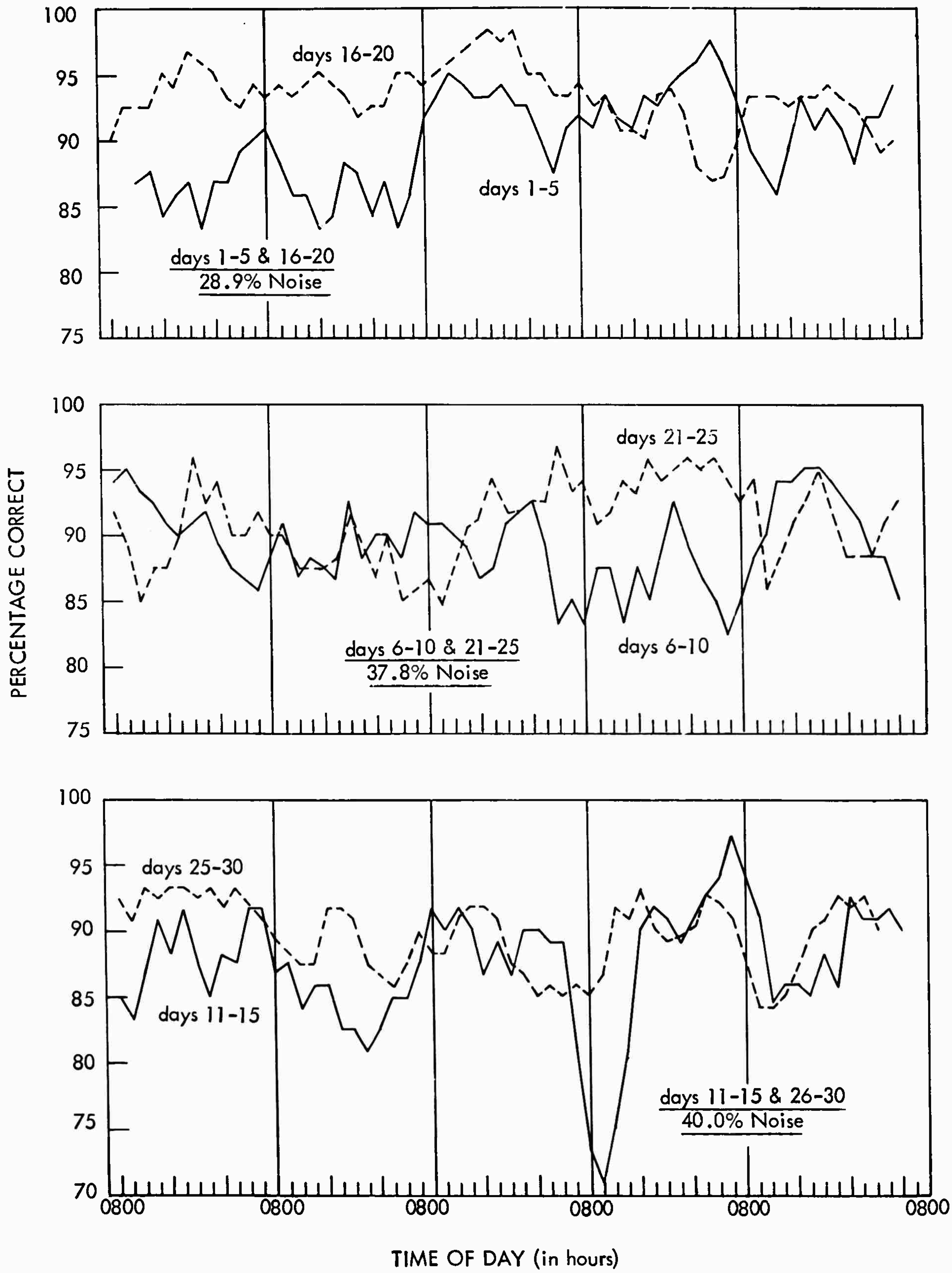

Figure 27. Mean percentage of correct commander's final decisions in target identification without simultaneous presentation of code-lock problems: HOPE-III.

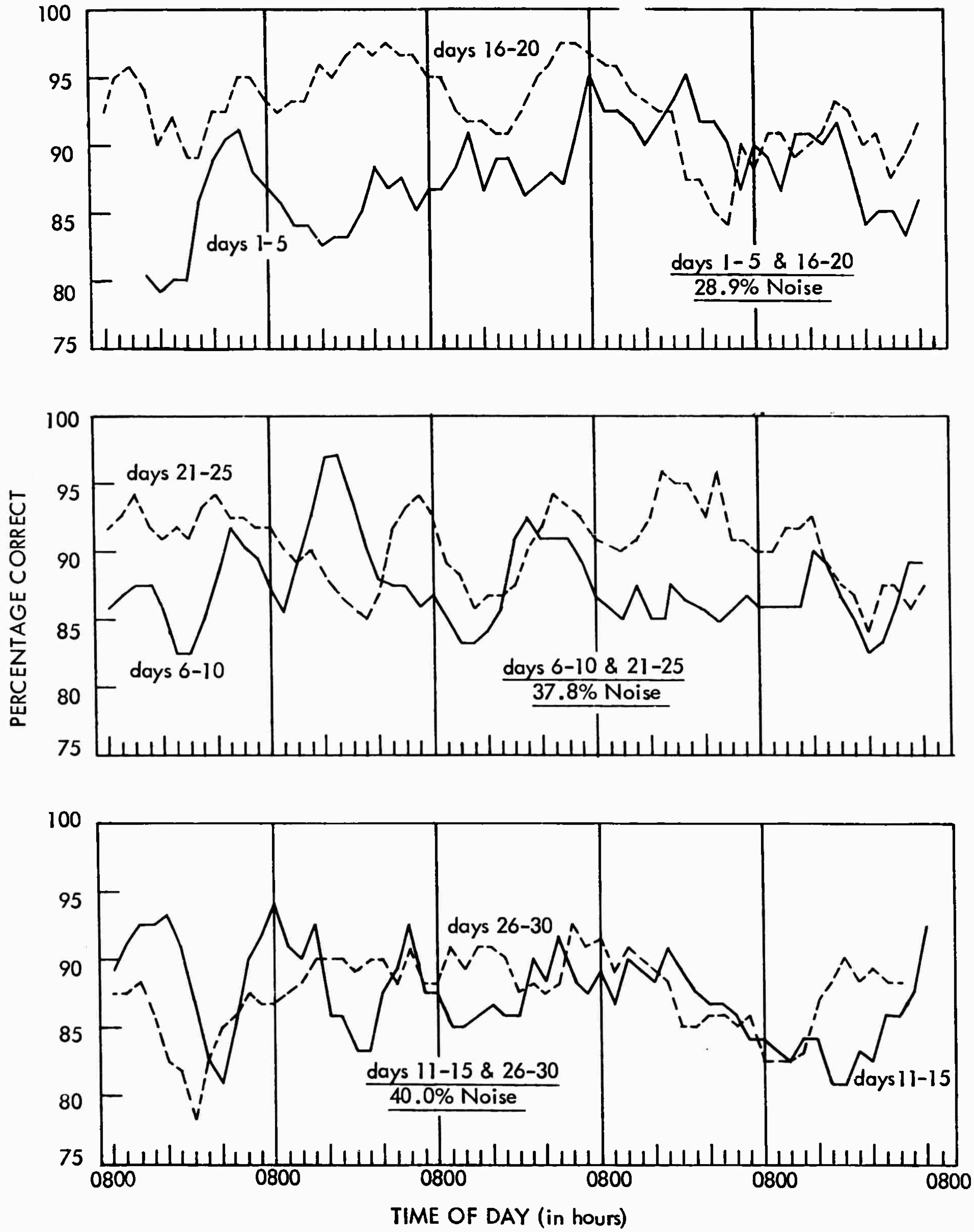

Figure 28. Mean percentage of correct commander's final decisions in target identification with simultaneous presentation of code-lock problems: HOPE-III.

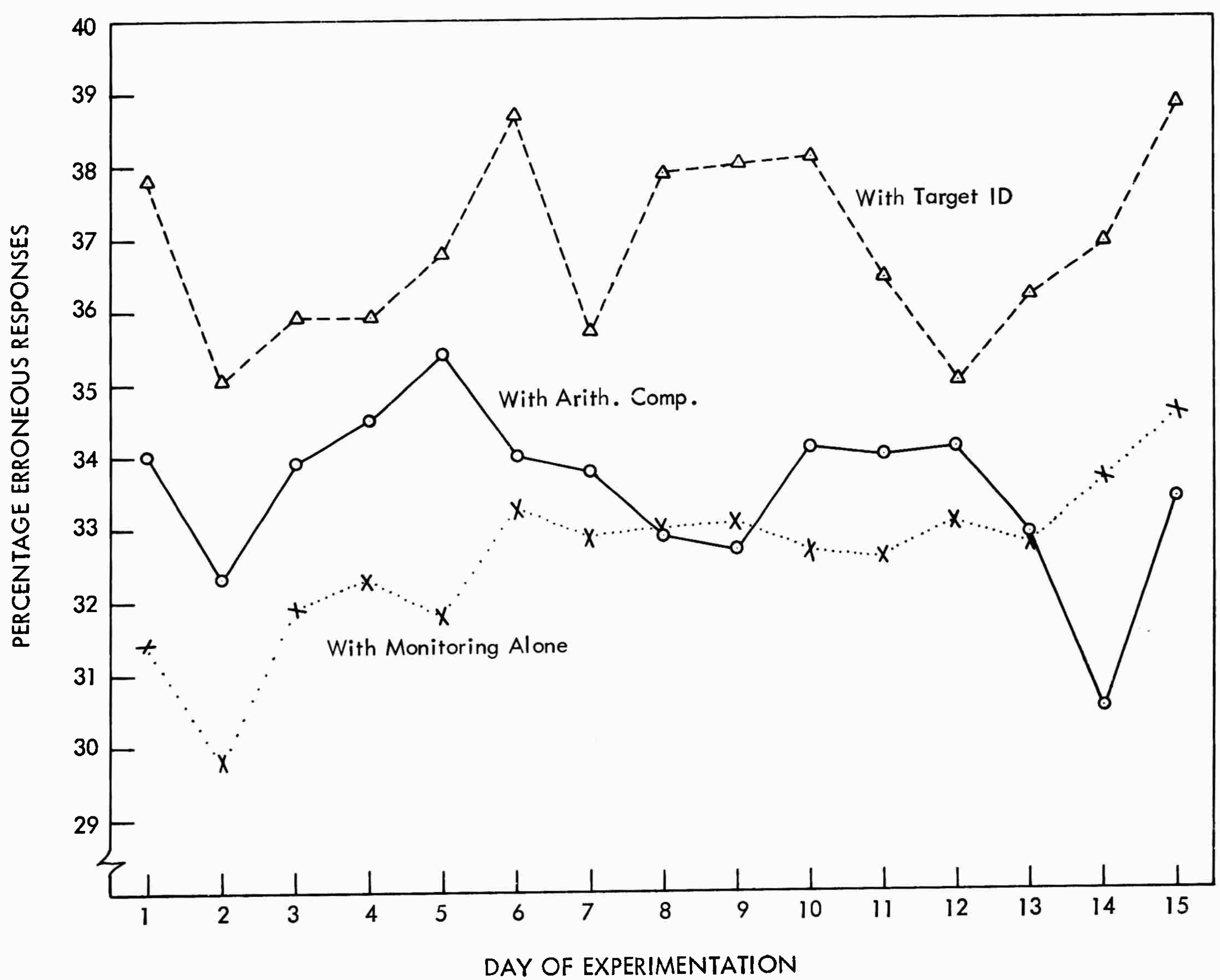

Figure 29. Mean percentage of erroneous code-lock responses under three response conditions: HOPE-II.

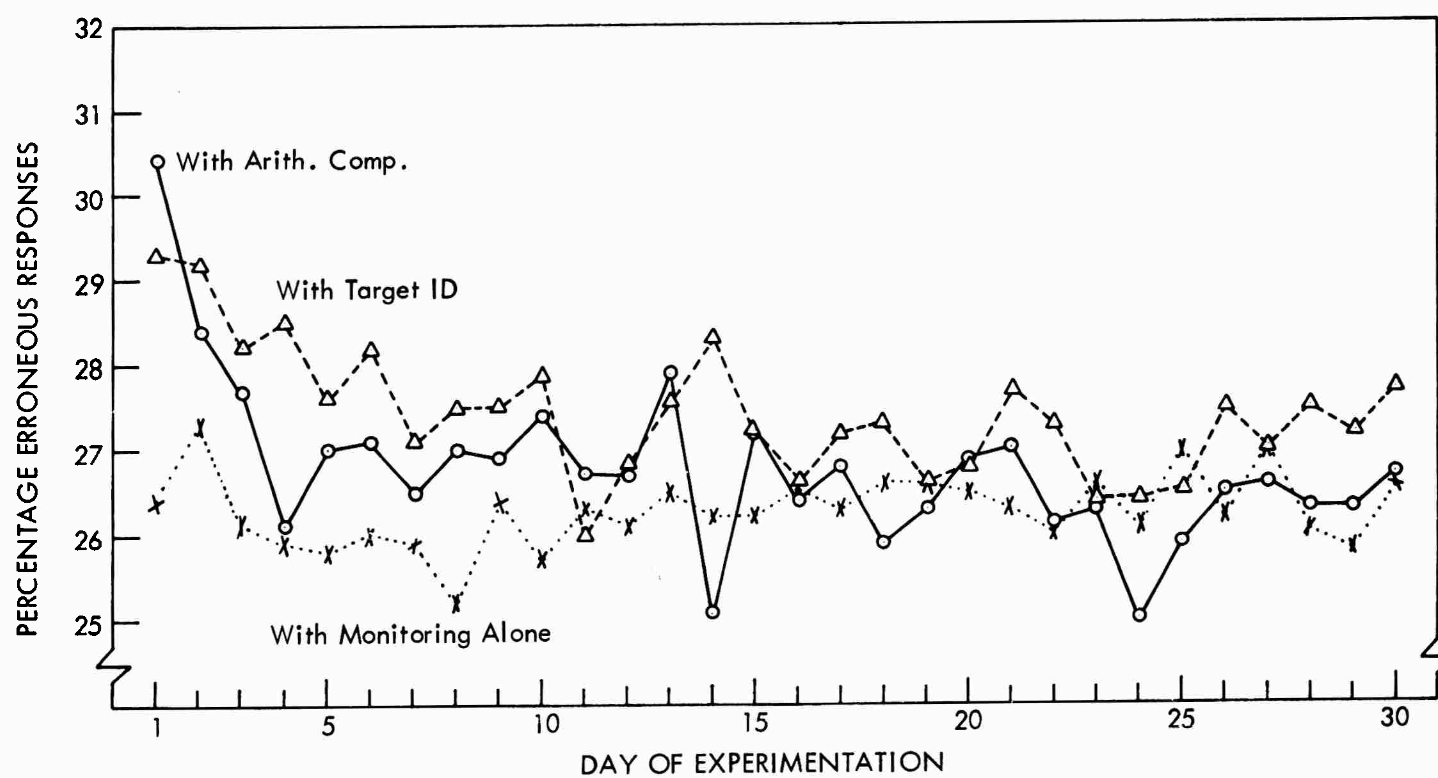

Figure 30. Mean percentage of erroneous code-lock responses under three response conditions: HOPE-III.

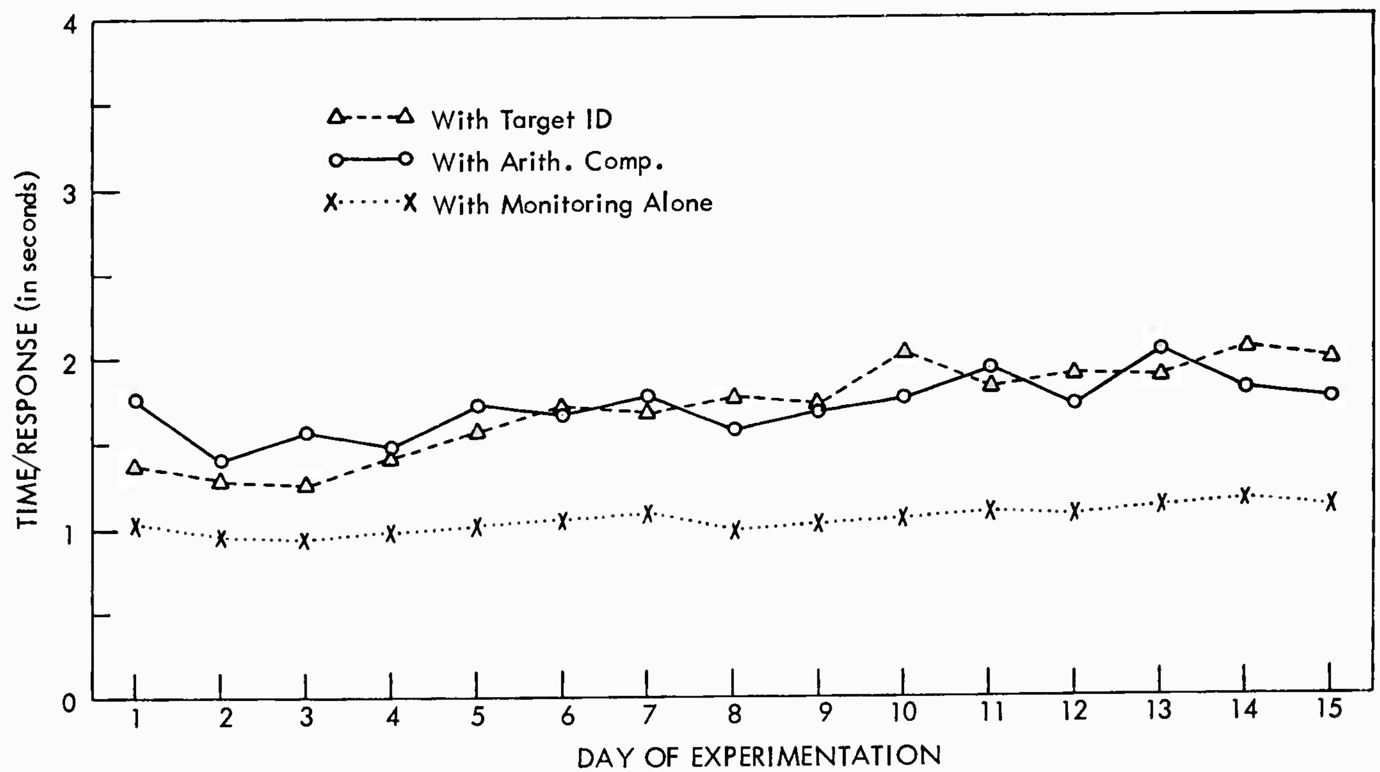

Figure 31. Mean time per code-lock response under three response conditions: HOPE-II.

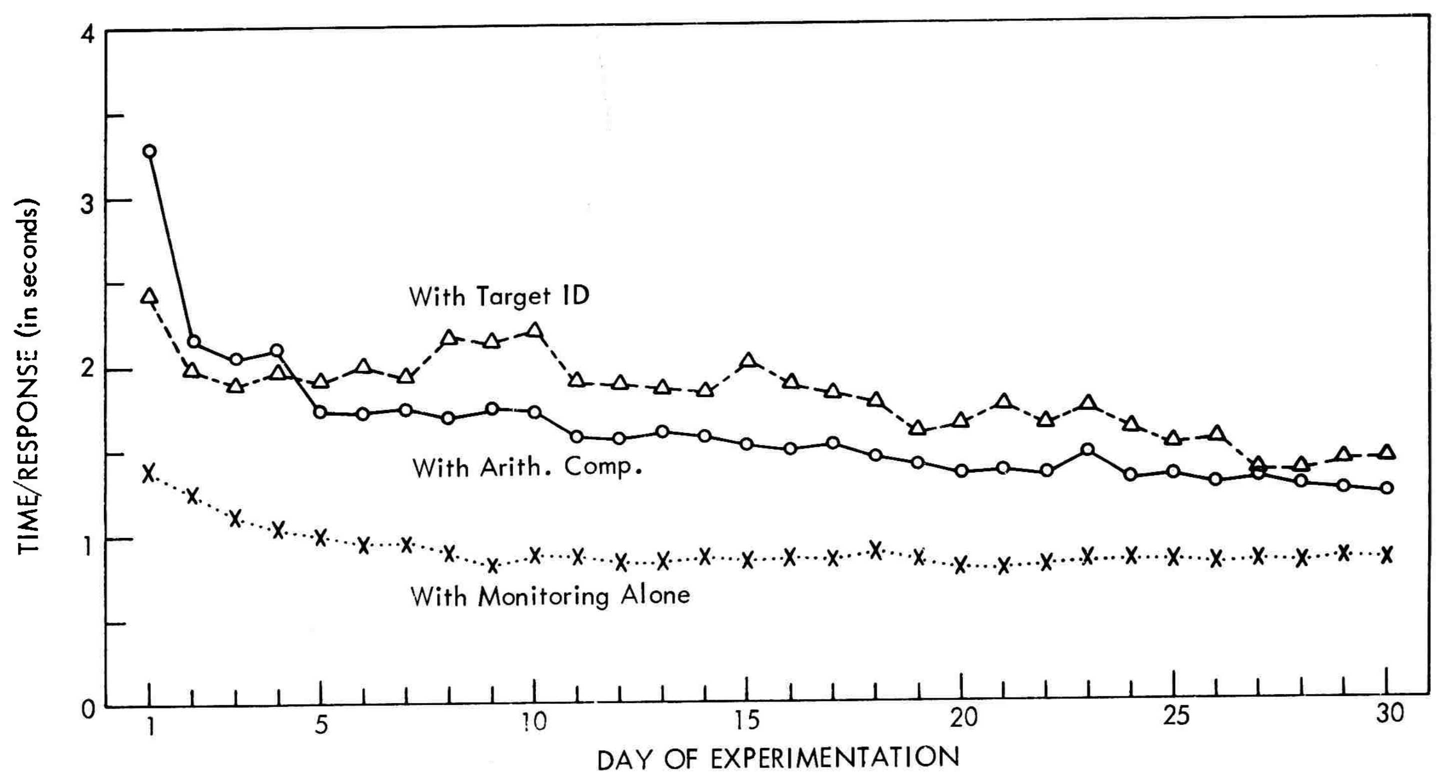

Figure 32. Mean time per code-lock response under three response conditions: HOPE-III.

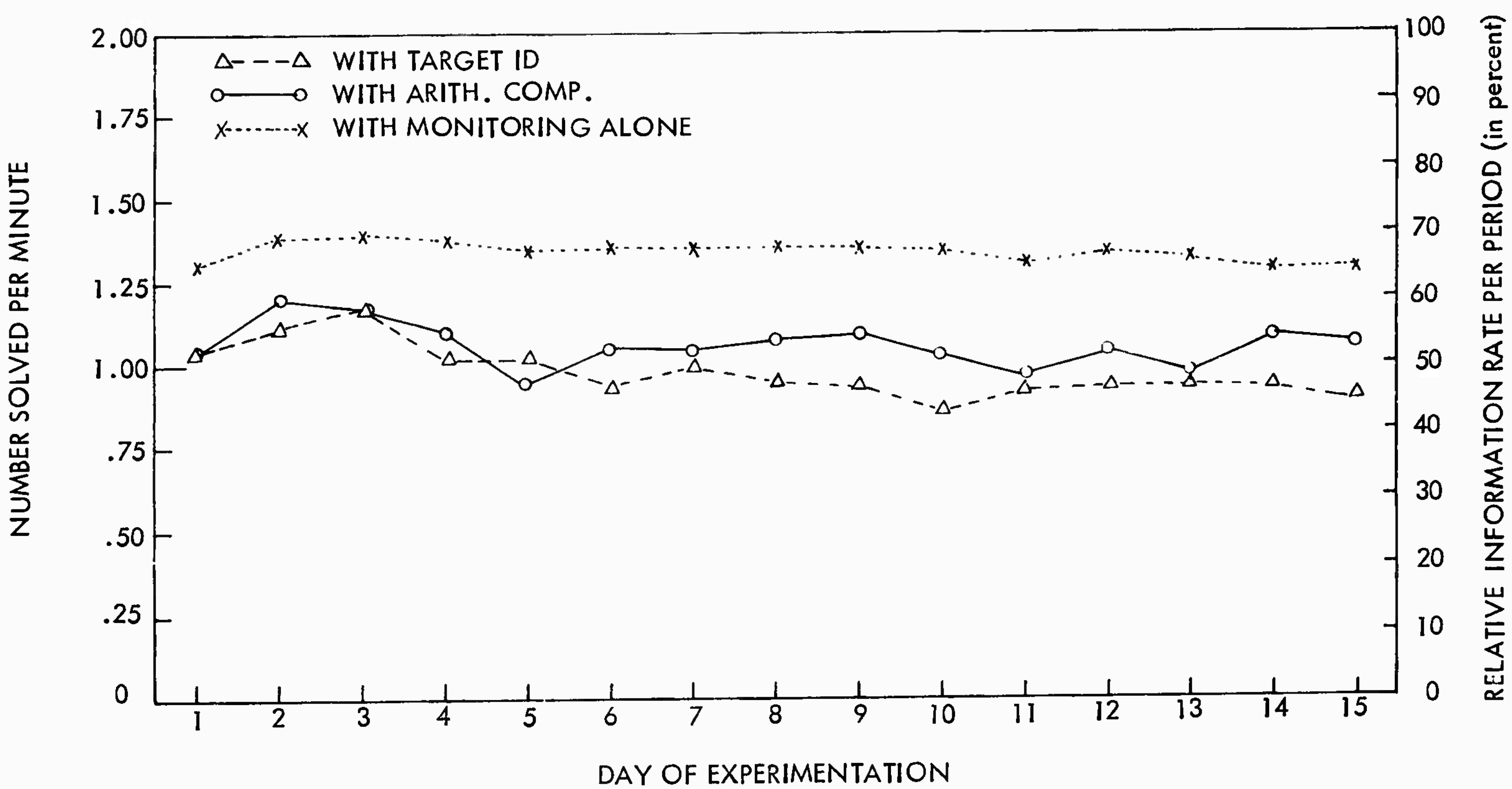

Figure 33. Mean number of code-lock problems solved per minute under three response conditions: HOPE-II. Also (scale on right), relative information rate per period (in percentage of maximum attainable rate).

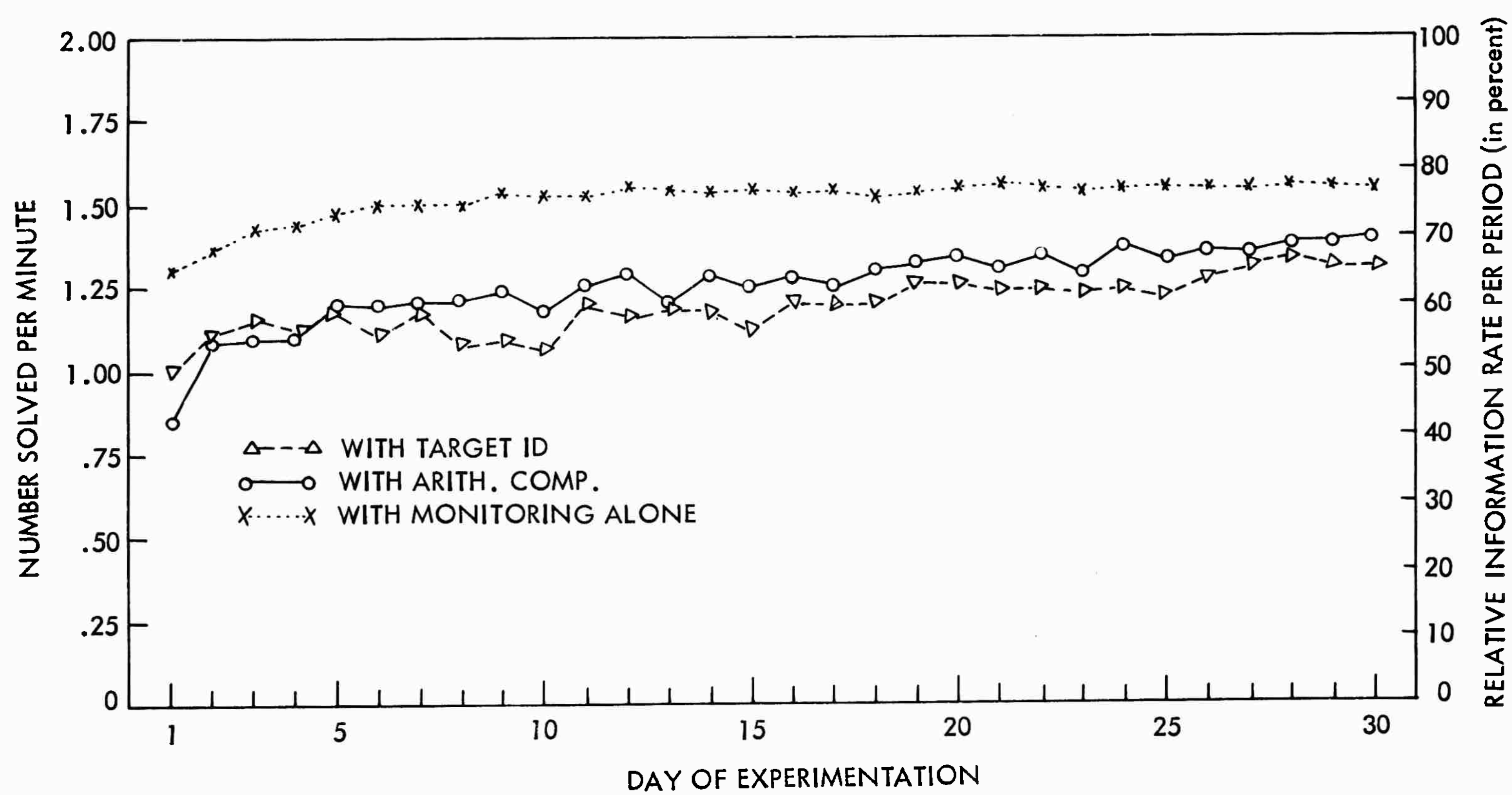

Figure 34. Mean number of code-lock problems solved per minute under three response conditions: HOPE III. Also (scale on right), relative information rate per period (in percentage of maximum attainable rate).